Joint Modeling and Estimation for Recurrent Events, Longitudinal Measurements and Survival Data

by
Qing Cai

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March, 2017

Abstract

Recurrent events together with longitudinal measurements are commonly observed in follow-up studies where the observation is terminated by censoring or a primary failure event. In this dissertation, we developed joint modeling methods for recurrent events, longitudinal measurements and survival data. We developed a joint model where the dependence of longitudinal measurements, recurrent event process and time to failure event is modeled through rescaling the time index. The general idea is that the trajectories of all biology processes of subjects in the survivors' population are elongated or shortened by the rate identified from a model for the failure event. To avoid making disputing assumptions on recurrent events or biomarkers after the failure event (such as death), the model is constructed on the basis of survivors' population. The model also possesses a specific feature that, by aligning failure events as time origins, the backward-in-time model of recurrent events and longitudinal measurements shares the same parameter values with the forward time model. The proposed method can be generalized to analyze left-truncated data. We also developed methods to model the terminal behaviors of stochastic processes and estimate the backward change point which is identified based on the failure event in backward time index. As a particular application of our methods in the studies of biomarkers' predictability in relation to disease, the rate of change of biomarkers or longitudinal measurements could start to shift, accelerate or decelerate at a specific time prior to the occurrence of a primary failure event, and our model is aimed to estimate this special time point termed as a backward change point. Modeling the backward change point is challenging because longitudinal measurements, recurrent events and failure event are usually correlated, and the observed data are incomplete due to censoring of follow-up. Our methods are on the basis of incomplete data and allow the

join distribution of failure event, recurrent events and longitudinal measurements to be unspecified except for a few common assumptions. The general idea is that we model the longitudinal measurement process retrospectively starting from the failure event and the failure time will be a condition of the model just like a covariate. The statistical properties, simulation studies and real data examples are conducted.

Adviser: Professor Mei-Cheng Wang

Acknowledgements

I would like to express my wholehearted thanks to those who helped me and walked with me during my five-year doctoral study at the Department of Biostatistics, Johns Hopkings University. My research would not be possible without them. First and foremost, I would like to sincerely thank my advisor, Dr. Mei-Cheng Wang, for her supreme guidance, encouragement and surport throughout my study and life over these years. I have benefited a lot from her critical thinking, extensive knowledge, and great passion in research. She always gave me many insights and led me to the right direction in research. I have been most fortunate to have her as my advisor and work with her. I am also sincerely grateful to Dr. Karen Bandeen-Roche, Dr. Brain Caffo, Dr. Ani Eloyan and other faculty members here for their continuous contributions to make this graduate programs outstanding. I also would like to thank Mary Joy, Jiong and other staff members for their kind help and support in these years. I would also like to thank Dr. Marilyn Albert, Dr. Anja Soldan and other members in BIOCARD study group for their support, and it was an valuable experience for me to work with them as a research assistant.

I would also like to give my best wishes and sincere thanks to Tianchen, David, Yuting, Dan, Jiawei, Emily and other students in this department for their help and friendship. I will always remember the time when we studied together and it is a wonderful memory for me. Thanks to Yifei for always providing help and anwsering my stupid questions with patients. Thanks to Yi and Daisy for their help and support when we worked together as research assistants. Thanks to Juemin, Shaojie, Huitong, Chen and others for their help for my job search.

Most importantly, I would like to express my deepest appreciation to my family members and friends who always care about me and support me. Thanks to my dear

parents Zhonghai Cai and Ruiqun Xue for their endless love. I would never have been possible to reach this milestone without their constant encouragement and support. My love and gratitude to them are far beyond words.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview of Statistical Problems

Recurrent events together with longitudinal measurements are frequently encountered in follow-up studies. In biomedical applications, two types of longitudinal measurements are commonly observed: (i) repeated measurements collected at sampling times, and (ii) marker measurements observed when recurrent events occur. In case (i), longitudinal measurements are assumed to be a stochastic process which exists continuously over time, such as CD4 cell counts in HIV studies or other disease-related biomarkers, where recurrent events are sampling times. In case (ii), typical examples are studies with repeated marker measurements observed upon the occurrence of recurrent events, where both recurrent events and marker measurements are of scientific interest; an example is the medical charge upon the occurrence of hospitalization. For either case, the observation of longitudinal measurements and recurrent events could be terminated by censoring or a primary failure event such as death. Despite the difference in the data generating mechanisms, these two kinds of data share the same notations.

An example of type (i) longitudinal measurements can be the data from the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) cohort study which aims at identifying biomarkers associated with the development of Alzheimer's Disease (AD) procession. The study was administrated by NIH from 1995 to 2005,

and was re-established by a research group at Johns Hopkins School of Medicine after being stopped for four years. Subjects enrolled in the study were cognitively normal at baseline and data including cognitive performance testing scores were collected annually or per half year during the study.

For the type (ii) longitudinal measurement, the data from the clinical trial conducted by Terry Beirn Community Programs for Clinical Research on AIDS is a good illustration where the longitudinal measurements were only observed with the occurance of a recurrent event process. The study was federally funded and conducted by national network of community-based research groups. The study compared two kinds of treatments, didanosine (ddI) and zalcitabine (ddC), for HIV-infected patients who previously had failed treatment with or were intolerant to zidovudine. (Abrams et al., 1994[1]) In the trial, 230 patients were randomly sellected to receive ddI treatment and 237 to receive ddC. Patients were followed to death as the primary endpoint, during which totally 463 opportunistic infections were observed (172 in the ddI group and 191 in the ddc group). For each opportunistic infection, a severity score was provided by doctors and HIV-infected patients to indicate the disease progression (Neaton et al., 1994[28]).

Analyzing this type of data is challenging because of the complex association within the data structure. Firstly, ignoring the dependence of longitudinal measurements to survival data leads to biased results. As a simple explanation, the failure times are possibly correlated to the longitudinal measurements and the survivors' population will change with the time going. Thus, the changes of the longitudinal measures are because of the time factor, the covariate effects and also the influence of the survivors' population variation. The traditional and commonly used methods for longitudinal data analyses such as the generalize estimation equation models (Liang and Zeger, 1986[19]) and the mixed effect models (Laird and Ware, 1982[18]) can not be applied in this scenarios because these models were developed on the basis

of independent failure event and constant population. Additionally, these two types of models assume the sampling processes are independent with the longitudinal data such that they can not be used to analyze the type (ii) longitudinal measurements even without informative failure events.

Besides modeling the longitudinal data in forward time order, it is also of interest to study the terminal behavior of stochastical processes in backward time index in studies of disease progression. Particularly, it is often of practical meaning to study biomarker performance prior to the occurrence of failure events by aligning failure events as time origins and counting time backward. For example, in studies of Alzheimer's Disease the rate of change in biomarker measurement before diagnosis of disease is widely recognized as an important index for predicting the disease (Hall et al., 2000[8]; Wilson et al., 2007[43]). However, the conventional forward perspective of stochastic processes are not designed for the terminal behavior of processes.

Additionally, when studying trajectories of biomarkers or other longitudinal measurements before the failure event such as death, researchers sometimes find that the rate of change of biomarkers start to shift, accelerate or decelerate, at some special time point which is prior to the failure event by a gap time. We name this special type of change point as backward change point to distinguish it from the traditional forward change point which occurs after the origin point by a constant period of time. Backward change point is of scientific interest for biology procession study and disease diagnosis. In particular, for Alzheimer's disease (AD) studies, Jack et al. (2010)[12] hypothesized that the decline of biomarkers began to accelerate prior to diagnosis in order such that different biomarkers characterized the disease during different stages. Knowing the acceleration order of biomarkers decline will help to measure disease progression precisely, and then therapeutic intervention can be given to patients properly.

Consider the model where the failure time is T and the decline of biomarker

begins to accelerate at time point $T - d$ for those subjects with failure time longer than d, where d is a constant gap time. Ideally, if all of the subjects are followed until diagnosis, the backward change point model can be analyzed by aligning the biomarker measurements retrospectively with diagnosis as the time origin, and then use techniques for forward change point models (e.g., Slate and Cronin, 1997[32]; Skates, Pauler, and Jacobs ,2001[31]) to estimate the parameter d.

However, in most follow-up studies, subjects may be censored due to design limitations, early drop out or other reasons. For censored subjects, not only the time of failure event is missing, the longitudinal measurements between censoring and failure event are also unobserved. As a result, the new time origins of censored data cannot be set and the observed longitudinal measurements cannot be indexed in backward time index. Thus, the censoring problem of longitudinal and survival analysis studies in backward time index is more complex than that of the studies in forward time index. However, as a character of time-to-event data, disregarding the censoring problem will lead to biased results even if the cenoring is noninformative.

In this dissertation, we have developed novel methods for joint modeling recurrent events, longitudinal measurements and survival data which allow the association within the data structure. We have also conducted research to model the terminal behavior of stochastic processes and especially estimate of backward change point.

1.2 Organization

The dissertation is organized as follows. We have conducted the literature reviews for related statistical methods for joint modeling of recurent events, longitudinal measurements and survival data, and the change point estimation for backward stochastic processes in the Chapter 2. Chapter 3 contains our methods for joint modeling of recurent events, longitudinal measurements and survival data, and the simulation studies and real data illustrations are also included. In the Chapter 4, we present

our methods for the backward change poiont estimation, provide the asymptotical inference results, and illustrate our methods by simulation studies and the real data exsample. Our thoughts and discussions are in the Chapter 5.

Chapter 2

Literature Review

2.1 Literature Review of Joint Modeling of Recurrent Events, Longitudinal Measurements and Survival Data

Recurrent events together with longitudinal measurements are commonly observed in medical or public health follow-up studies. As mentioned in the Chapter 1, there are two types of longitudinal measurements commonly observed: (i) repeated measurements collected at sampling times (e.g. such as CD4 cell counts in HIV studies or other disease-related biomarkers), and (ii) marker measurements observed when recurrent events occur (e.g. the medical charge upon the occurrence of hospitalization). For either case, the observation of longitudinal measurements and recurrent events can be terminated by censoring or a primary failure event such as death. Despite the difference in the data generating mechanisms, these two kinds of data share the same notations.

In the absence of informative sampling times, many authors (Wulfsohn and Tsiatis, 1997[45]; Henderson et al., 2000[9]; Xu and Zeger, 2001[47]; Song, Davidian, and Tsiatis, 2002[33]; Vonesh, Greene, and Schluchter, 2006[40]; Song and Wang, 2008[34]) used the shared frailty model to analyze longitudinal measurements and time-to-event data, where multiple layers of models with shared underling random variables are created for different data components. The general idea of shared frailty models is that

different data components are assumed to be independent of each other conditioning on the shared random variables, and hence the correlation within the data structure is explained by the shared random variables. However, since these models assume that the sampling times are indepdent witht the failure time and the longitudinal measurements, none of them can be applied to longitudinal measurements of type (ii).

Informative sampling times is another important topic in longitudinal studies. For example, medical cost or some biomarkers would be measured at each time of hospitalization. The process of hospitalization can be viewed as a recurrent event process which could carry information for longitudinal variables. Thus, joint modeling longitudinal measurements and recurrent events data will be beneficial and even necessary in some situations. Lin and Ying (2001)[22] proposed semiparametric models for longitudinal data with irregular sampling times, where the longitudinal response variable is assumed to be independent of the recurrent event process conditioning on covariates. Lin, Scharfstein and Rosenheck (2004)[23] relaxed the assumption of independent-relationship between longitudinal variables and observational times to some degree. The method of Sun et al. (2005)[35] allowed a flexible correlation between the two components of the data, where the recurrent event process is assumed to be Poisson process. Liang, Lu and Ying (2009)[20] developed a shared random effects model, where random effects were assumed to have a specific relationship defined by a link function. In all these methods, assumptions of unrelated terminal events are necessary.

Abundant attempts have been adopted to jointly model longitudinal measurements, recurrent events, and time to failure events data. The shared frailty models (Sun, Sun, and Liu, 2007[36]; Liu, Huang, and O'Quigley, 2008[25]; Liu and Huang, 2009[24]; Kim et al., 2012[14]; Sun et al., 2012[37]) form a popular approach as well. However, some of these models can not be adopted in the situation of randomly sam-

pled measurements, i.e., type (i), and many have limitations in computation complexity. In addition, various approaches have been proposed for modeling recurrent events together with informative failure event jointly (Wang, Qin, and Chiang, 2001[42]; Huang and Wang, 2004[10]; Liu, Wolfe, and Huang, 2004[26]; Ye, Kalbfleisch, and Schaubel, 2007[48]; Zeng and Cai, 2010[50]). Most of these approaches consider the association within the data structure by using the shared frailty method, except the models in Sun et al. (2005)[35].

2.2 Literature Review of Backward Change Point Estimation

In studies of disease progression, it is often of interest to study biomarker performance prior to the occurrence of failure events by aligning failure events as time origins and counting time backward. For example, in studies of Alzheimer's Disease the rate of change in biomarker measurement before diagnosis of disease is widely recognized as an important index for predicting the disease (Hall et al., 2000[8]; Wilson et al., 2007[43]). However, the conventional forward perspective of stochastic processes are not designed for the terminal behavior of processes. To our knowledge, Chan and Wang (2010)[4] was the first to study the backward stochastic processes and develop a nonparametric estimation method where the failture events were treated as time origins and the time was counted in backward order. Chan and Wang (2016) [3] further developed a three layer semiparametric regression model for jointly modeling the survival data, the backward recurrent event proess and the marker measured at occurance of backeard recurrent events.

Additionally, when studying trajectories of biomarkers or other longitudinal measurements before the failure event such as death, researchers sometimes find that the rate of change of biomarkers start to shift, accelerate or decelerate, at some special time point which is prior to the failure event by a gap time. We name this special

type of change point as backward change point to distinguish it from the traditional forward change point which occurs after the origin point by a constant period of time. Backward change point is of scientific interest for biology procession study and disease diagnosis. In particular, for Alzheimer's disease (AD) studies, Jack et al. (2010)[12] hypothesized that the decline of biomarkers began to accelerate prior to diagnosis in order such that different biomarkers characterized the disease during different stages. Knowing the acceleration order of biomarkers decline will help to measure disease progression precisely, and then therapeutic intervention can be given to patients properly.

Consider the model where the failure time is T and the decline of biomarker begins to accelerate at time point $T - d$ for those subjects with failure time longer than d, where d is a constant gap time. Ideally, if all of the subjects are followed until diagnosis, the backward change point model can be analyzed by aligning the biomarker measurements retrospectively with diagnosis as the time origin, and then use techniques for forward change point models (e.g., Slate and Cronin, 1997[32]; Skates, Pauler, and Jacobs ,2001[31]) to estimate the parameter d.

However, in most follow-up studies, subjects may be censored due to design limitations, early drop out or other reasons. For censored subjects, not only the time of failure event is missing, the longitudinal measurements between censoring and failure event are also unobserved. As a result, the new time origins of censored data cannot be set and the observed longitudinal measurements cannot be indexed in backward time index. Thus, the censoring problem of longitudinal and survival analysis studies in backword time index is more complex than that of the studies in forward time index. However, as a character of time-to-event data, disregarding the censoring problem will lead to biased results even if the cenoring is non-informative. To our knowledge, there are few valid methods of modeling backward change point problems. Some researchers (Hall et al., 2000[8]; Wilson et al., 2011[43]) have already found the

scientific meaning of backward change point in applications, but did not consider the censoring issue in their models. Additionally, longitudinal measurements and the failure event are usually correlated and the association within data structure should also be considered. Unfortunately, the current methods of joint modeling of lingitudinal measuremetns and time-to-event data (Wulfsohn and Tsiatis, 1997[45]; Henderson et al., 2000[9]; Xu and Zeger, 2001[47]; Song et al., 2002[33]; Vonesh et al., 2006[40]; Song and Wang, 2008[34]) do not consider the backward change point problem.

Chapter 3

Joint Modeling of Longitudinal, Recurrent Events and Failure Time Data for Survivors Population

3.1 Introduction

Recurrent events together with longitudinal measurements are frequently encountered in follow-up studies. In biomedical applications, two types of longitudinal measurements are commonly observed: (i) repeated measurements collected at sampling times, and (ii) marker measurements observed when recurrent events occur. In case (i), longitudinal measurements are assumed to be a stochastic process which exists continuously over time, such as CD4 cell counts in HIV studies or other disease-related biomarkers, where recurrent events are sampling times. In case (ii), typical examples are studies with repeated marker measurements observed upon the occurrence of recurrent events, where both recurrent events and marker measurements are of scientific interests; an example is the medical charge upon the occurrence of hospitalization. For either case, the observation of longitudinal measurements and recurrent events could be terminated by censoring or a primary failure event such as death. Despite the difference in the data generating mechanisms, these two kinds of data share the same notations. This chapter presents a semiparametric joint model framework for both types of longitudinal measurements, (i) and (ii), with covariates information.

In the absence of informative sampling times, many authors (Wulfsohn and Tsiatis, 1997[45]; Henderson et al., 2000[9]; Xu and Zeger, 2001[47]; Song, Davidian, and Tsiatis, 2002[33]; Vonesh, Greene, and Schluchter, 2006[40]; Song and Wang, 2008[34]) used the shared frailty model to analyze longitudinal measurements and time-to-event data, where multiple layers of models with shared underlining random variables are created for different data components. The general idea of shared frailty models is that different data components are assumed to be independent of each other conditioning on the shared random variables, and hence the correlation within the data structure is explained by the shared random variables. However, since these models assume that the sampling times are indepdent, none of them can be applied to longitudinal measurements of type (ii). Abundant attempts have been adopted to jointly model longitudinal measurements, recurrent events, and time to failure events data. The shared frailty models (Sun, Sun, and Liu, 2007[36]; Liu, Huang, and O'Quigley, 2008[25]; Liu and Huang, 2009[24]; Kim et al., 2012[14]; Sun et al., 2012[37]) form a popular approach as well. However, some of these models may not be applicable to randomly sampled measurements, i.e., type (i) data, due to specific model restrictions for recurrent event processes. In addition, various approaches have been proposed for jointly modeling longitudinal measurements and recurrent events with uncorrelated failure event (Lin and Ying, 2001[22]; Lin, Scharfstein, and Rosenheck, 2004[23]; Sun et al., 2005[35]; Liang, Lu, and Ying, 2009[20]), and modeling recurrent events together with informative failure event jointly (Wang, Qin, and Chiang, 2001[42]; Huang and Wang, 2004[10]; Liu, Wolfe, and Huang, 2004[26]; Ye, Kalbfleisch, and Schaubel, 2007[48]; Zeng and Cai, 2010[50]). Most of these approaches consider the data correlation by using the shared frailty method, except the models in Lin and Ying (2001)[22], Lin et al. (2004)[23] and Sun et al. (2005)[35].

In this chapter, we consider a joint model of longitudinal measurements, recurrent events and time to failure events, which is a useful alternative to the shared frailty

method and is applicable to both two types of longitudinal measurements (i) and (ii). The time rescaling technique was previously adopted by Ghosh and Lin (2003)[7] and Huang and Wang (2003)[11] for modeling recurrent events together with informative failure events. Luo, Wang, and Huang (2008)[46] also provides a detailed comparison of different models. Our model has the features that the trajectories of all biology processes are targeted on subjects in the survivors' population, where a subject's failure time is elongated or shortened by the rate identified from a failure time model for the failure event. Hence, our method can be developed without disputing assumption on the recurrent events or biomarkers after death. In contrast with the shared frailty method, our model does not involve latent variable and the proposed methodological procedure is computationally simpler.

Besides the forward time model, by aligning failure events as time origins, we will also study statistical model and inference of recurrent events and longitudinal measurements backward in time. For many diseases, including Alzheimer's Disease, the changes in biomarker performance before diagnosis of disease is widely recognized as an important index for predicting the disease (Hall et al., 2000[8]; Wilson et al., 2007[43]). The conventional stochastic process models are always forward in time and not designed to study terminal behavior of processes. However, backward process models are more relevant for studying these terminal behavior (Chan and Wang, 2010[4]; Chan and Wang, 2016). In this chapter, we will show that our model, though developed in forward time scale, has the consistent interpretation in the backward models under proper assumptions.

In Section 3.4, the method is generalized to the case where left truncation in failure time data is present. It handles the sampling designs when the study recruits only subjects who have experienced the initiating event but have not experienced the failure event. Under left truncation, it is well known that the time from initiating event to failure event tends to be longer than the failure time from the target pop-

ulation. We will use a data example from the Alzheimer's Disease to illustrate the proposed method.

The chapter is organized as follows. Section 3.2.1 introduces the joint model and discusses the interpretation of the joint model. An alternative model is presented in Section 3.2.2 and we show that the model has consistent interpretation in backward time perspective. Section 3.3 describes the estimation procedures and develops statistical properties of the proposed semiparametric estimators. The generalization to left truncation is discussed in Section 3.4. In Section 3.5, we evaluate the performance of the estimating methods by simulation studies. Real data from an Alzheimer's Disease study and an AIDS study are used as illustrating examples for data analysis in Section 3.6. Additional discussions are in included Section 3.7 to conclude the chapter.

3.2 Survivors' Model

3.2.1 A Time-Adjusted Forward Model

Let T be the time from an initial event to the failure event which is in continuous time scale, $\boldsymbol{Z}$ a $p \times 1$ vector of covariates, and $R^*(t)$ the counting process of the number of sampling times at or before time t. The longitudinal process $Y(t)$ is measured repeatedly at time t where $dR^*(t) = 1$. Suppose that the study is conducted in the time interval $[0, \tau]$ and, potentially, the data information of $\{T, R^*(t), Y(t)\}$ is observed for $0 \leq t \leq \tau$. In reality, due to limitation in experimental design or other reasons, the data are subject to right censoring and only partially observed. Let C_R be the censoring time for observing the sampling times and the longitudinal measurements $\{R^*(\cdot), Y(\cdot)\}$. Let C_T be the censoring time for the failure event. When framing the proposed model, we shall consider only the recurrent events and the longitudinal measurements occurring before the failure event, i.e. $t \leq T$.

Consider the following assumption:

(A1) $\{T, R^*(\cdot), Y(\cdot)\}$ is independent of (C_T, C_R) conditioning on $\boldsymbol{Z}$.

Under Assumptions (A1), we consider a time-adjusted forward model to characterize the joint relationship of $\{T, R^*(\cdot), Y(\cdot)\}$: At any $t \geq 0$, assume

$$\begin{cases} \ln T = \boldsymbol{\beta}_T' \boldsymbol{Z} + \ln T_0, \quad \text{where } \boldsymbol{Z} \text{ is independent of } T_0 \\ \mathrm{E}\left\{ R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) \mid T \geq t e^{\boldsymbol{\beta}_T' \boldsymbol{Z}}, \boldsymbol{Z} \right\} = e^{\boldsymbol{\beta}_R' \boldsymbol{Z}} \mathrm{E}\left\{ R_0^*(dt) \mid T_0 \geq t \right\}, \\ \mathrm{E}\left\{ Y\left(t e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \right) \mid T \geq t e^{\boldsymbol{\beta}_T' \boldsymbol{Z}}, R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = 1, \boldsymbol{Z} \right\} = \mathrm{E}\left\{ Y_0(t) \mid T_0 \geq t, R_0^*(dt) = 1 \right\} + \boldsymbol{\beta}_Y' \boldsymbol{Z}, \end{cases} \tag{3.1}$$

where $\boldsymbol{\beta_T}$, $\boldsymbol{\beta_R}$ and $\boldsymbol{\beta_Y}$ is a $p \times 1$ vector of parameters, and T_0, $R_0^*(\cdot)$ and $Y_0(\cdot)$ are respectively the baseline failure time, the baseline recurrent event process and the baseline longitudinal process. Here the baseline variables for subjects with $\boldsymbol{Z} = \mathbf{0}$. The three components $\{T_0, R_0^*(\cdot), Y_0(\cdot)\}$ are distribution-free and possibly correlated. Of note, the notations $\{T_0, R_0^*(\cdot), Y_0(\cdot)\}$ represent individual-level variables or processes, and $\{T_{0i}, R_{0i}^*(\cdot), Y_{0i}(\cdot)\}$ will be used in later sections as variables or processes for subject i. In Model (3.1), the notation $R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right)$ is used to indicate the jumps of the recurrent event process on the adjusted time scale, i.e. $R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} t \right) - R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} t^- \right)$.

We use Figure 3.1 to introduce the main idea of time rescaling. Suppose there are longitudinal measurements for treatment and controll groups until their failure events. We assume that their longitudinal measurements will have the same type of trajectory since we suppose the disease development process is similar for the whole population. Assume the speed of disease development process of the treatment group is slow, and therefore the survival time is longer and the data trajectory is elongated compared to that of the control group. If we adjust the data trajectory of the treatment group and let it has the same disease development speed with that of control group, the two biomarker trajectories should be the same except for an additional shift. The time rescaling rate is defined by the ratio of survival times. In summary, our method supposes that one part of the covariate effect is time rescaling effect, and another part is represented by the additive shift term.

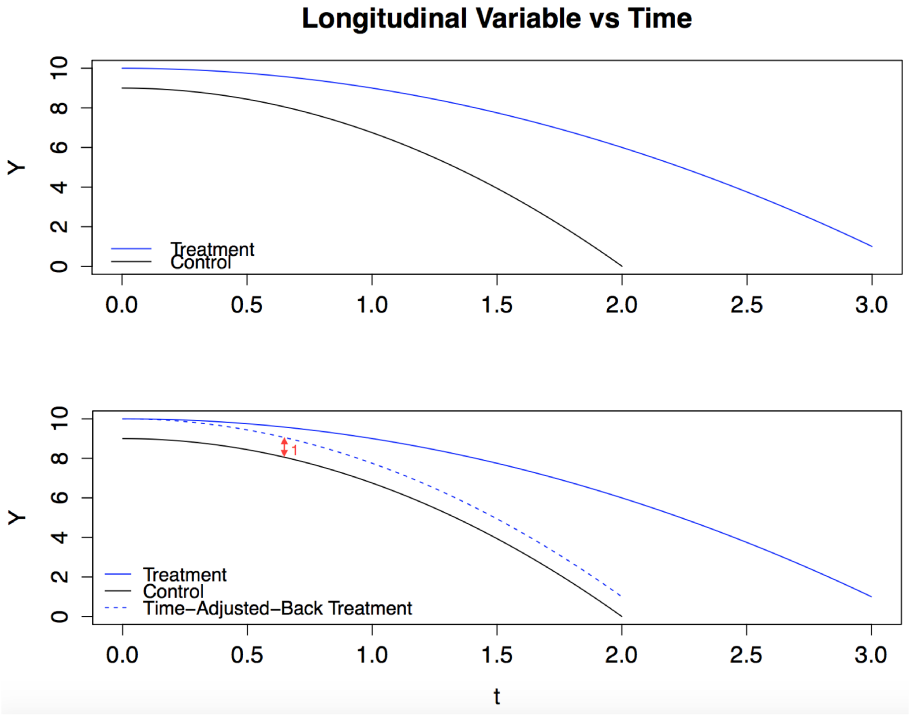

Figure 3.1: The Time Rescaling Method

Model (3.1) has an important feature as a *survivors' model*, in which only those who have not experienced the failure event are included when modeling the recurrent events and longitudinal measurements. The model assumes that the trajectories of a subject's recurrent event and longitudinal measurement processes in the survivors' population are elongated or shortened by the regression function identified from the accelerated failure time model (Kalbfleisch and Prentice, 1980). More precisely, the time scale in $\{R^*(\cdot), Y(\cdot)\}$ is transformed from $[0, T_0]$ to $[0, T]$ using $e^{\boldsymbol{\beta}_T' \boldsymbol{Z}}$ as the multiplicative adjustment. Besides the covariate effect on time-rescaling, covariates are also allowed to influence the rate function of $R^*(\cdot)$ by a multiplicative term $\exp(\boldsymbol{\beta}_R' \boldsymbol{Z})$ and influence the trajectory function $Y(\cdot)$ by an additive term $\boldsymbol{\beta}_Y' \boldsymbol{Z}$, respectively.

Note that Assumption (A1) specifies the independent censoring assumption. The condition $(T \geq t)$ in Model (3.1) has the practical interpretation that $\{R^*(\cdot), Y(\cdot)\}$ is required to exist only for the time before the failure event, and therefore allows Model

(3.1) to capture the covariate effects on the survivors' population and avoids the arguable issue of longitudinal variables or recurrent events after the failure time. Additionally, the survivors' population changes when t varies, and this special characteristic is indicated by the condition $(T \geq t)$. The conditioning event $R^*\left(\exp(\boldsymbol{\beta}_T'\boldsymbol{Z})dt\right) = 1$ in the third layer of Model (3.1) is a natural condition for type (ii) data, since the marker measurement $Y\left(t\exp(\boldsymbol{\beta}_T'\boldsymbol{Z})\right)$ exists and is observed given the occurrence of a recurrent event. For type (i) longitudinal measurement, Model (3.1) models the mean trajectory and covariate effect only for the observed $Y(.)$, but the result generalizes to the underlying $Y(.)$ if $Y(.)$ and $R(.)$ are independent given Z. Similar conditions are employed for type (i) data in Lin and Ying (2001)[22].

As mentioned in Section 1, Model (3.1) is applicable to the two types of longitudinal measurements (i) and (ii), where the rate of sampling process in the rescaled time index is changed by a multiplier $\exp(\boldsymbol{\beta}_T'\boldsymbol{Z})$ compared to baseline. For the special case when the rate of sampling points is pre-planned and constant overtime, the rate of $R^*(t)$ is not influenced by covariates and we have $\boldsymbol{\beta}_R = \boldsymbol{\beta}_T$. In this case, the non-zero $\boldsymbol{\beta}_R$ is the rescaling effect rather than the covariate effect. Note that the case $\boldsymbol{\beta}_R = \mathbf{0}$ implies that the recurrent event process is affected by the covariates only through rescaling the time to failure event, where the recurrent event process is stretched or compressed. Thus, a positive or negative $\boldsymbol{\beta}_R$ reflects the additional inflation or deflation of the frequency of $R^*(\cdot)$ explained by the covariates.

3.2.2 An Alternative Model and Its Backward Property

By rescaling time index, we construct an alternative model, termed as Model (3.2). This model is slightly stronger than Model (3.1), and possesses an attractive feature of having consistent interpretation in both the forward and backward models as will be explained later. Under the independent censoring assumption (A1), Model (3.2)

assumes

$$\begin{cases} \ln T = \boldsymbol{\beta}_T' \boldsymbol{Z} + \ln T_0, \quad \text{where } \boldsymbol{Z} \text{ is independent of } T_0 \\ \mathrm{E}\left\{ R^*\left(te^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \right) \mid T, \boldsymbol{Z} \right\} = e^{\boldsymbol{\beta}_R' \boldsymbol{Z}} \mathrm{E}\left\{ R_0^*(t) \mid T_0 \right\}, \quad t \in [0, T_0] \\ \mathrm{E}\left\{ Y\left(te^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \right) \mid T, R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = 1, \boldsymbol{Z} \right\} = \mathrm{E}\left\{ Y_0(t) \mid T_0, R_0^*(dt) = 1 \right\} + \boldsymbol{\beta}_Y' \boldsymbol{Z}, \ t \in [0, T_0] \end{cases} \tag{3.2}$$

Model (3.2) extends the joint model of failure time and recurrence events studied by Huang and Wang (2003) to a joint model with the additional longitudinal measurement $Y(\cdot)$. Note that by conditioning on T, the baseline failure time T_0 plays a role similar to a subjective-specific random effect or frailty, and the model identifies the regression effect on $\boldsymbol{Z}$. Huang and Wang (2003) indicated that their recurrent event model extends to

$$\mathrm{E}\left\{ R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) \mid T \geq te^{\boldsymbol{\beta}_T' \boldsymbol{Z}}, \boldsymbol{Z} \right\} = e^{\boldsymbol{\beta}_R' \boldsymbol{Z}} \mathrm{E}\left\{ R_0^*(dt) \mid T_0 \geq t \right\}.$$

In our case, under Model (3.2), the longitudinal measurement satisfies

$$\begin{aligned} & \mathrm{E}\left\{ Y\left(te^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \right) \mid T \geq te^{\boldsymbol{\beta}_T' \boldsymbol{Z}}, R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = 1, \boldsymbol{Z} \right\} \\ =& \mathrm{E}\left[\mathbb{1}_{\left[T \geq te^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \right]} \mathrm{E}\left\{ Y\left(te^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \right) \mid T, R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = 1, \boldsymbol{Z} \right\} \mid R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = 1, \boldsymbol{Z} \right] \\ & \left\{ \mathbb{P}\left(T \geq te^{\boldsymbol{\beta}_T' \boldsymbol{Z}} \mid R^*\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt \right) = 1, \boldsymbol{Z} \right) \right\}^{-1} \\ =& \mathrm{E}\left[\mathbb{1}_{[T_0 \geq t]} \left\{ \mathrm{E}\left(Y_0(t) \mid T_0, R_0^*(dt) = 1 \right) + \boldsymbol{\beta}_Y' \boldsymbol{Z} \right\} \mid R_0^*(dt) = 1, \boldsymbol{Z} \right] \\ & \left\{ \mathbb{P}\left(T_0 \geq t \mid R_0^*(dt) = 1, \boldsymbol{Z} \right) \right\}^{-1} \\ =& \mathrm{E}\left\{ Y_0(t) \mid T_0 \geq t, R_0^*(dt) = 1 \right\} + \boldsymbol{\beta}_Y' \boldsymbol{Z}. \end{aligned}$$

Thus, the regression model of $Y(\cdot)$ in Model (3.1) would be validated by Model (3.2), and a similar argument extends to the model for $R^*(\cdot)$ to conclude that the validity of Model (3.2) implies Model (3.1). Essentially, Model (3.1) can be thought of as a marginal model of Model (3.2), since T is used as a part of conditional statistics in the latter but not in the former.

Interestingly, as a special property, Model (3.2) implies the validity of a backward time model with the same parameter value. The backward time model can be used

to study the terminal behaviors of the biological processes where the failure event is aligned as the time origin and time is counted backward. Backward time models offer a natural and direct way to study terminal behavior of recurrent markers. Using the human immunodeficiency virus (HIV) infection as an illustrating example, it is scientifically and clinically interesting to understand the pattern of the frequency and severity of opportunistic infections before death. Evidence suggests that HIV-infected patients experienced higher frequency of AIDS-defining events before death, where the frequencies could vary with gender, risk behaviors or geographic location (Chan et al., 1995[2]). Similar terminal behaviors are studied for Alzheimer's Disease (Wilson et al., 2007[43]), renal disease (Usvyat et al., 2013[38]) and functional decline in the general population (Lunney et al., 2003[27]). To set notation, let the backward time index be denoted by t^B which stands for the time counted retrospectively from the failure event. Define the backward process of the longitudinal measurement as $Y^B(t^B;T) = Y(T - t^B)$ and the backward recurrent event process as $R^B(t^B;T) = R^*(T) - R^*((T - t^B)^-)$, where $0 \le t^B \le T$.

Under Assumptions (A1), by aligning failure events as time origins and counting time backward, it can be proved that Model (3.2) holds if and only if the following backward model holds: For any $t \in [0, T_0]$,

$$\begin{cases} \ln T = \boldsymbol{\beta}_T' \boldsymbol{Z} + \ln T_0, \quad \text{where } \boldsymbol{Z} \text{ is independent of } T_0 \\ \mathrm{E}\left\{ R^B\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt^B; T \right) \mid T, \boldsymbol{Z} \right\} = e^{\boldsymbol{\beta}_R' \boldsymbol{Z}} \mathrm{E}\left\{ R_0^B\left(dt^B; T_0 \right) \mid T_0 \right\}, \\ \mathrm{E}\left\{ Y^B\left(t^B e^{\boldsymbol{\beta}_T' \boldsymbol{Z}}; T \right) \mid T, R^B\left(e^{\boldsymbol{\beta}_T' \boldsymbol{Z}} dt^B; T \right) = 1, \boldsymbol{Z} \right\} = \\ \qquad\qquad \mathrm{E}\left\{ Y_0^B\left(t^B; T_0 \right) \mid T_0, R_0^B\left(dt^B; T_0 \right) = 1 \right\} + \boldsymbol{\beta}_Y' \boldsymbol{Z}. \end{cases} \tag{3.3}$$

Model (3.3) assumes that the trajectories of a subject's backward recurrent event and longitudinal measurement processes are elongated or shortened by the scale identified from the accelerated failure time model, where covariates influence the rate of $R^B(\cdot)$ by a multiplicative term $\exp(\boldsymbol{\beta}_R' \boldsymbol{Z})$ and influence the measurement $Y^B(\cdot)$ by an additive term $\boldsymbol{\beta}_Y' \boldsymbol{Z}$, respectively. Model (3.3) is consistent with Model (3.2) in the sense that

they are equivalent to each other and share the same parameter values. This property implies that estimation of Model (3.3) can be derived through estimation of Model (3.2), as what we will establish in the next section. The proof for equivalence between Models (3.2) and (3.3) is provided in Section 3.8.1.

3.3 Estimation

In this section we introduce and discuss the estimation procedures of Models (3.1) and (3.2). Define $X_T = T \wedge C_T$, $X_R = T \wedge C_R$, and $R(t) = R^*(t \wedge X_R)$, where $a \wedge b = \min(a, b)$. Let $\Delta = \mathbb{1}_{[T \leq C_T]}$ be the censoring indicator of the failure event. Assume the observations $\{X_{Ti}, \Delta_i, X_{Ri}, R_i(\cdot), Y_i(\cdot), \boldsymbol{Z}_i\}$, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.). For most applications, the censoring time for recurrent events and longitudinal measurements is either the same as or less than the censoring time for the failure event. We therefore assume $\mathbb{P}(C_R \leq C_T) = 1$.

We construct three-layer estimation functions for $\boldsymbol{\beta}_T$, $\boldsymbol{\beta}_R$, $\boldsymbol{\beta}_Y$, separately but in a certain ordering, for Model (3.1). As the first step, the weighted log-rank estimating equations (Tsiatis, 1990) is popular among other approaches (Buckley and James, 1979; Ritov, 1990) for estimating $\boldsymbol{\beta}_T$ in the AFT model. We define the adjusted risk-set indicator of failure event by $\xi_{Ti}(t; \boldsymbol{\beta}_T) = \mathbb{1}_{\left[X_{Ti} \geq t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i)\right]}$ and the adjusted counting process of failure event by $N_{Ti}(t; \boldsymbol{\beta}_T) = \mathbb{1}_{\left[X_{Ti} \leq t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i)\right]} \Delta_i$. The estimating function is

$$\boldsymbol{U}_T(\boldsymbol{\beta}_T) = \sum_{i=1}^{n} \int_0^{\infty} W_T(t; \boldsymbol{\beta}_T) \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} \xi_{Tj}(t; \boldsymbol{\beta}_T) \boldsymbol{Z}_j}{\sum_{j=1}^{n} \xi_{Tj}(t; \boldsymbol{\beta}_T)} \right\} dN_{Ti}(t; \boldsymbol{\beta}_T), \tag{3.4}$$

where $W_T(t; \boldsymbol{\beta}_T)$ is a non-negative weight function. A zero-crossing of $\boldsymbol{U}_T(\boldsymbol{\beta}_T) = \boldsymbol{0}$ exists, termed as $\hat{\boldsymbol{\beta}}_T$, as the estimator of $\boldsymbol{\beta}_T$ which is strongly consistent and asymptotically normal under regular conditions (Ying, 1993[49]).

In the next step we use the approach of Huang and Wang (2003) to construct an estimator for $\boldsymbol{\beta}_R$. For given $\boldsymbol{\beta}_T$, define the adjusted risk-set indicator of recurrent

event process by $\xi_{Ri}(t;\boldsymbol{\beta}_T) = \mathbb{1}_{\left[X_{Ri} \geq t\exp(\boldsymbol{\beta}_T'\boldsymbol{Z}_i)\right]}$ and the adjusted counting process of recurrent event process as $N_{Ri}(t;\boldsymbol{\beta}_T) = R_i(t\exp(\boldsymbol{\beta}_T'\boldsymbol{Z}_i)) = R_i^*(t\exp(\boldsymbol{\beta}_T'\boldsymbol{Z}_i) \wedge X_{Ri})$. The estimating function for $\boldsymbol{\beta}_R$ is

$$\boldsymbol{U}_R(\boldsymbol{\beta}_R;\boldsymbol{\beta}_T) = \sum_{i=1}^{n}\int_0^{\infty} W_R(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)\left\{\boldsymbol{Z}_i - \tilde{\boldsymbol{Z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)\right\} dN_{Ri}(t;\boldsymbol{\beta}_T), \tag{3.5}$$

where $\tilde{\boldsymbol{Z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R) = \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\xi_{Rj}(t;\boldsymbol{\beta}_T)e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}{\sum_{j=1}^{n}\xi_{Rj}(t;\boldsymbol{\beta}_T)e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}$ and $W_R(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ is a non-negative weight function. If $W_R(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ does not depend on $\boldsymbol{\beta}_R$, $\boldsymbol{U}_R(\boldsymbol{\beta}_R;\boldsymbol{\beta}_T)$ is a monotone function of $\boldsymbol{\beta}_R$ and its zero-crossing is a consistent estimator of $\boldsymbol{\beta}_R$ (Huang and Wang, 2003). Replacing $\boldsymbol{\beta}_T$ by $\hat{\boldsymbol{\beta}}_T$, we solve the equation $\boldsymbol{U}_R(\boldsymbol{\beta}_R;\hat{\boldsymbol{\beta}}_T) = \mathbf{0}$ instead and denote the zero-crossing solution as $\hat{\boldsymbol{\beta}}_R$.

In the third step of estimation, we estimate the parameters $\boldsymbol{\beta}_Y$ via an estimation equation of $(\boldsymbol{\beta_Y}, \hat{\boldsymbol{\beta}}_{\boldsymbol{T}}, \hat{\boldsymbol{\beta}}_{\boldsymbol{R}})$. The rate function of recurrent events process in survivors' population at baseline is $\mathrm{E}\left\{dR_0^*(t) \mid T_0 \geq t\right\} = \mathbb{P}\left(T_0 \geq t\right)^{-1} d\mathrm{E}\left\{R_0^*(t)\right\}$, and the baseline cumulative recurrence rate function of survivors' population is $\Lambda_{R_0}(t) = \int_0^t \mathbb{P}\left(T_0 \geq u\right)^{-1} d\mathrm{E}\left\{R_0^*(u)\right\}$. To simplify notation, we define

$$\alpha_0(t) = \mathrm{E}\left\{Y_0(t) \mid T_0 \geq t, R_0^*\left(dt\right) = 1\right\}$$

which is an unspecified function of t, and define

$$\mathscr{A}_0(t) = \int_0^t \mathrm{E}\left\{Y_0(t)R_0^*(du) \mid T_0 \geq u\right\} = \int_0^t \alpha_0(u)d\Lambda_{R_0}(u) = \int_0^t \frac{\alpha_0(u)}{\mathbb{P}\left\{T_0 \geq u\right\}} d\mathrm{E}\left\{R_0^*(u)\right\}. \tag{3.6}$$

All the baseline information is structurally combined in $\mathscr{A}_0(t)$, because one can prove that

$$\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right) - \boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\} dN_{Ri}(t;\boldsymbol{\beta}_T) \mid \xi_{Ri}(t;\boldsymbol{\beta}_T)\right] = \xi_{Ri}(t;\boldsymbol{\beta}_T)e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_i} d\mathscr{A}_0(t). \tag{3.7}$$

Therefore, a mean-zero stochastic process for the ith subject can be expressed as

$$\begin{aligned} &M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0) \\ &= \int_0^t \left\{Y_i\left(ue^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right) - \boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\} dN_{Ri}(u;\boldsymbol{\beta}_T) - \int_0^t \xi_{Ri}(u;\boldsymbol{\beta}_T)e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_i} d\mathscr{A}_0(u), \end{aligned}$$

which leads to two estimation equations:

$$\sum_{i=0}^{n} M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0)=0, \tag{3.8}$$

$$\sum_{i=0}^{n}\int_0^{\infty} W_Y(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y)\boldsymbol{Z}_i\; dM_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0)=\mathbf{0}. \tag{3.9}$$

Based on (3.8), we develop an estimator for $d\mathscr{A}_0(t)$ and $\mathscr{A}_0(t)$, respectively, as

$$\begin{aligned} d\hat{\mathscr{A}}_0(t) &= \frac{\sum_{i=1}^{n}\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}dN_{Ri}(t;\boldsymbol{\beta}_T)}{\sum_{j=1}^{n}\xi_{Rj}(t;\boldsymbol{\beta}_T)e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}},\\ \hat{\mathscr{A}}_0(t) &= \sum_{i=1}^{n}\int_0^{t}\frac{Y_i\left(ue^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i}{\sum_{j=1}^{n}\xi_{Rj}(u;\boldsymbol{\beta}_T)e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}dN_{Ri}(u;\boldsymbol{\beta}_T). \end{aligned} \tag{3.10}$$

Replacing $\mathscr{A}_0(t)$ by $\hat{\mathscr{A}}_0(t)$ in (3.9), we create an estimating function of $\boldsymbol{\beta}_Y$ with given $\{\boldsymbol{\beta}_T,\boldsymbol{\beta}_R\}$:

$$\begin{aligned} \boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R) =& \sum_{i=1}^{n}\int_0^{\infty} W_Y(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y)\left\{\boldsymbol{Z}_i-\tilde{\boldsymbol{Z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)\right\}\\ & \left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}dN_{Ri}(t;\boldsymbol{\beta}_T). \end{aligned} \tag{3.11}$$

Simple algebraic calculation yields

$$\begin{aligned} &\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)\\ =&\sum_{i=1}^{n}\int_0^{\infty} W_Y(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y)\left(\boldsymbol{Z}_i-\tilde{\boldsymbol{Z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)\right)dM_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0). \end{aligned}$$

So, by similar arguments of Lin et al. (2000), $n^{-1/2}\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ will converge weakly to a zero mean Gaussian process with continuous sample paths. A zero-root of $\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)=\mathbf{0}$ will be a consistent estimator of $\boldsymbol{\beta}_Y$.

Replacing $\boldsymbol{\beta}_T$ and $\boldsymbol{\beta}_R$ by $\hat{\boldsymbol{\beta}}_T$ and $\hat{\boldsymbol{\beta}}_R$ respectively, we estimate $\boldsymbol{\beta}_Y$ by the zero-root solution of the equation $\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\hat{\boldsymbol{\beta}}_T,\hat{\boldsymbol{\beta}}_R)=\mathbf{0}$. If the weight function $W_Y(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y)$ does not depend on $\boldsymbol{\beta}_Y$ and

$$\left\{\sum_{i=1}^{n}\int_0^{\infty} W_Y\left(t;\hat{\boldsymbol{\beta}}_T,\hat{\boldsymbol{\beta}}_R\right)\left(\boldsymbol{Z}_i-\tilde{\boldsymbol{Z}}\left(t;\hat{\boldsymbol{\beta}}_T,\hat{\boldsymbol{\beta}}_R\right)\right)\boldsymbol{Z}_i' dN_{Ri}\left(t;\hat{\boldsymbol{\beta}}_T\right)\right\}$$

is invertible, there is a closed form of the solution as

$$\hat{\boldsymbol{\beta}}_Y = \left\{ \sum_{i=1}^{n} \int_0^{\infty} W_Y\left(t; \hat{\boldsymbol{\beta}}_T, \hat{\boldsymbol{\beta}}_R\right) \left(\boldsymbol{Z}_i - \tilde{\boldsymbol{Z}}\left(t; \hat{\boldsymbol{\beta}}_T, \hat{\boldsymbol{\beta}}_R\right)\right) \boldsymbol{Z}_i' dN_{Ri}\left(t; \hat{\boldsymbol{\beta}}_T\right) \right\}^{-1}$$
$$\left\{ \sum_{i=1}^{n} \int_0^{\infty} W_Y\left(t; \hat{\boldsymbol{\beta}}_T, \hat{\boldsymbol{\beta}}_R\right) \left(\boldsymbol{Z}_i - \tilde{\boldsymbol{Z}}\left(t; \hat{\boldsymbol{\beta}}_T, \hat{\boldsymbol{\beta}}_R\right)\right) Y_i\left(te^{\hat{\boldsymbol{\beta}}_T' \boldsymbol{Z}_i}\right) dN_{Ri}\left(t; \hat{\boldsymbol{\beta}}_T\right) \right\},$$

which substantially relaxes the burden of computation.

Under Model (3.2), the estimation procedures are essentially the same as those under Model (3.1) because Model (3.1) is induced from Model (3.2): By re-defining

$$d\mathscr{A}_0(t) = \mathrm{E}\left[Y_0(t) R_0^*(dt) \mid T_0 \geq t, R_0^*(dt) = 1\right]$$

the regression parameters can be estimated by solving estimating equations (3.4), (3.5) and (3.11) (see Section 3.8.2). The estimation of the backward model, (3.3), can also be achieved via Model (3.1).

We develop asymptotic properties of the regression estimators with constant weight functions, which are shared by Models (3.1) and (3.2). Define $\boldsymbol{\beta} = (\boldsymbol{\beta}_T', \boldsymbol{\beta}_R', \boldsymbol{\beta}_Y')'$, $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_T', \hat{\boldsymbol{\beta}}_R', \hat{\boldsymbol{\beta}}_Y')'$, and the trivariate estimating function

$$\boldsymbol{U}(\boldsymbol{\beta}) = (\boldsymbol{U}_T(\boldsymbol{\beta}_T)', \boldsymbol{U}_R(\boldsymbol{\beta}_R; \boldsymbol{\beta}_T)', \boldsymbol{U}_Y(\boldsymbol{\beta}_Y; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R)')'.$$

Similar to Ying (1993)[49] and Lin et al. (2000), we introduce the following conditions:

(C1) $\boldsymbol{\beta}$ is restricted in a compact set.

(C2) $\{\boldsymbol{Z}_i, T_i, R_i^*(\cdot), Y_i(\cdot)\}$, $i = 1, \ldots, n$ are i.i.d. with uniform bound.

(C3) The densities of T and $dR^*(t)$ and their first order differential functions are bounded.

(C4) The censoring time C_T and C_R have uniformly bounded densities, termed as g_T and g_R respectively, e.g. there is B_c such that $|g_T(t)| < B_c$ and $|g_R(t)| < B_c$ for all t.

(C5) $\mathrm{E}|\min\{\ln T - \boldsymbol{\beta}_T'\boldsymbol{Z}, C_T, C_R\}|^{\phi} < \infty$ for some $\phi > 0$.

Theorem 3.3.1 *Under conditions (C1-5), $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta})$ converges weakly to a multivariate normal random variable with mean zero and variance denoted as Σ, and $\boldsymbol{U}(\boldsymbol{\beta})$ is asymptotically linear in the sense that there exists a matrix $\boldsymbol{A}_n$ such that for every sequence $d_n > 0$ with $d_n \to 0$ in probability, we have*

$$\sup_{\|\boldsymbol{b}-\boldsymbol{\beta}\| \le d_n} \| \boldsymbol{U}(\boldsymbol{b}) - \boldsymbol{U}(\boldsymbol{\beta}) - \boldsymbol{A}_n n(\boldsymbol{b}-\boldsymbol{\beta}) \| / (\sqrt{n} + n \| \boldsymbol{b}-\boldsymbol{\beta} \|) = o_p(1).$$

If the eigenvalues of $\boldsymbol{A}_n$ are all bounded away from zero for all large enough n and $\boldsymbol{A}_n \to A$ where A is nonsingular, there exists a closed neighborhood $\mathcal{N}$ containing $\boldsymbol{\beta}$ as its interior point such that $\hat{\boldsymbol{\beta}}$ is strongly consistent and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges to $N(\mathbf{0}, A^{-1}\Sigma(A^{-1})')$ weakly in $\mathcal{N}$.

The proof of Theorem 3.3.1 is presented in the Section 3.8.3. Of note, the matrix A involves the true distributions of recurrent events, longitudinal measurements and censoring with covariates, and is hard to estimate in practice. For data applications we will use the Bootstrap approaches to estimate the confidence intervals.

3.4 Extension to Left Truncation

In biomedical studies, left-truncated sampling is commonly adopted, where only those subjects who have experienced the initiating event but have not experienced the failure event are recruited and followed until the occurrence of failure event or censoring; see Wang, Brookmeyer, and Jewell (1993)[41], among others. The failure time of interest is still the time from initiating event to failure event in the target population, and this type of sampling results in left-truncated and right-censored failure time data along with accompanied recurrent events and longitudinal measurements. For example, in Alzheimer's Disease studies, the failure time T is age at onset of symptoms of Mild Cognitive Impairment (MCI), and the study recruited and followed only those

subjects who were disease-free at baseline. The observed data sample then forms an example of left-truncated and right-censored data.

Our Model (3.1) and the subsequently developed inferential approach can be generalized to left-truncated and right-censored data under the conditional independent left truncation and right-censoring assumption, i.e. $\{T, R^*(\cdot), Y(\cdot)\}$ is independent of (C_T, C_R, L) where L is the left trucation time, given covariates $\boldsymbol{Z}$. Suppose the observations $\{L_i, X_{Ti}, \Delta_i, X_{Ri}, R_i(\cdot), Y_i(\cdot), \boldsymbol{Z}_i\}$, $i = 1, \ldots, n$, are i.i.d. subject to $L_i < X_{Ti} \wedge X_{Ri}$. To estimate Model (3.1) under the left-truncated and right-censored sampling, we re-construct the risk set indicators and the counting processes (Andersen et al., 2012). In particular, we define $\xi_{Ti}^L(t; \boldsymbol{\beta}_T) = \mathbb{1}_{\left[X_{Ti} \geq t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i) \geq L_i\right]}$, $N_{Ti}^L(t; \boldsymbol{\beta}_T) = \mathbb{1}_{\left[L_i \leq X_{Ti} \leq t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i)\right]} \Delta_i$, $\xi_{Ri}^L(t; \boldsymbol{\beta}_T) = \mathbb{1}_{\left[X_{Ri} \geq t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i) \geq L_i\right]}$, and $N_{Ri}^L(t; \boldsymbol{\beta}_T) = R_i^*(t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i) \wedge X_{Ri}) \mathbb{1}_{\left[L_i \leq t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i)\right]}$. Estimation procedures similar to those in Section 3 can be achieved using the following three-layer estimation functions:

$$
\begin{aligned}
\boldsymbol{U}_T^L(\boldsymbol{\beta}_T) &= \sum_{i=1}^n \int_0^\infty W_T(t; \boldsymbol{\beta}_T) \left\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^n \xi_{Tj}^L(t; \boldsymbol{\beta}_T) \boldsymbol{Z}_j}{\sum_{j=1}^n \xi_{Tj}^L(t; \boldsymbol{\beta}_T)} \right\} dN_{Ti}^L(t; \boldsymbol{\beta}_T), \\
\boldsymbol{U}_R^L(\boldsymbol{\beta}_R; \boldsymbol{\beta}_T) &= \sum_{i=1}^n \int_0^\infty W_R(t; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R) \left\{ \boldsymbol{Z}_i - \tilde{\boldsymbol{Z}}^L(t; ; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R) \right\} dN_{Ri}^L(t; \boldsymbol{\beta}_T), \\
\boldsymbol{U}_Y^L(\boldsymbol{\beta}_Y; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R) &= \sum_{i=1}^n \int_0^\infty W_Y(t; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R, \boldsymbol{\beta}_Y) \left\{ \boldsymbol{Z}_i - \tilde{\boldsymbol{Z}}^L(t; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R) \right\} \left\{ Y_i \left(t e^{\boldsymbol{\beta}_T' \boldsymbol{Z}_i} \right) \right. \\
&\quad \left. - \boldsymbol{\beta}_Y' \boldsymbol{Z}_i \right\} dN_{Ri}^L(t; \boldsymbol{\beta}_T),
\end{aligned}
$$

where $\tilde{\boldsymbol{Z}}^L(t; ; \boldsymbol{\beta}_T, \boldsymbol{\beta}_R) = \frac{\sum_{j=1}^n \boldsymbol{Z}_j \xi_{Rj}^L(t; \boldsymbol{\beta}_T) e^{\boldsymbol{\beta}_R' \boldsymbol{Z}_j}}{\sum_{j=1}^n \xi_{Rj}^L(t; \boldsymbol{\beta}_T) e^{\boldsymbol{\beta}_R' \boldsymbol{Z}_j}}$.

3.5 Simulation Studies

We conducted simulation studies to assess the finite sample performance of the proposed methods. For each simulation study, we repeatedly generated $1,000$ simulated data sets. In each data set, the ith individual's sample is generated by the following procedure:

- Generate the covariate vector $\boldsymbol{Z}_i$.

- Generate the potential baseline failure time T_{0i} independently with $\boldsymbol{Z}_i$. Since $T_{0i} > 0$, we generate a random variable V_i independent with $\boldsymbol{Z}_i$ and define $T_{0i} = \exp(V_i)$. Thus, conditioning on T_{0i} is equivalent to conditioning on V_i.

- Define the failure time as $T_i = \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i) T_{0i}$.

- Conditioning on $\{V_i, \boldsymbol{Z}_i\}$, generate the sampling time process $R_i^*(\cdot)$ on $[0, \tau]$ as a nonhomogeneous Poisson process with intensity

$$\lambda_i(t; \boldsymbol{Z}_i, V_i) = \exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_i) \exp(-\boldsymbol{\beta}_T' \boldsymbol{Z}_i) h(t \exp(-\boldsymbol{\beta}_T' \boldsymbol{Z}_i), V_i),$$

  where $h(\cdot, \cdot)$ is a prespecified positive function. Of note, this intensity function satisfies the requirement of Model (3.2) because $\mathrm{E}\{R_i^*(t \exp(\boldsymbol{\beta}_T' \boldsymbol{Z}_i)) | T_i, \boldsymbol{Z}_i\} = \exp(\boldsymbol{\beta}_R' \boldsymbol{Z}_i) \int_0^t h(u, V_i) du$ and $\mathrm{E}\{R_{0i}^*(t) | T_{0i}, \boldsymbol{Z}_i = \boldsymbol{0}\} = \int_0^t h(u, V_i) du$.

- Given V_i and $\boldsymbol{Z}_i$, at each sampling time t, generate the longitudinal variable by $Y_i(t) = \boldsymbol{\beta}_Y' \boldsymbol{Z}_i + a(t \exp(-\boldsymbol{\beta}_T' \boldsymbol{Z}_i)) + \varepsilon_i(t)$, where $\varepsilon_i(t)$ has normal distribution $N(\mu(V_i), \sigma^2(V_i))$. Here $a(\cdot)$, $\mu(\cdot)$ and $\sigma^2(\cdot)$ are prespecified functions.

- Generate censoring times $\{C_{Ti}, C_{Ri}\}$ which depend on the covariate $\boldsymbol{Z}_i$ but do not depend on V_i. Here, we set $C_{Ti} = C_{Ri} = C_i$.

In Table 3.1 all the specific distributions used to generate the simulated data are listed. The simulation procedure follows Model (3.2) exactly and is determined by the prespecified components: τ, distributions of $\{\boldsymbol{Z}_i, T_{0i}, C_i\}$, functions $\{h(\cdot, \cdot), \mu(\cdot), \sigma^2(\cdot), a(\cdot)\}$, and parameters $\{\boldsymbol{\beta}_T, \boldsymbol{\beta}_R, \boldsymbol{\beta}_Y\}$. The simulation procedure is also valid for Model (3.1). We examined several settings, with continuous and discrete covariates, different covariate dimensions, various associations within $\{T, R^*(\cdot), Y(\cdot)\}$, and different sample sizes. For the estimating procedure, we set weighting functions as $W_T(t; \boldsymbol{\beta}_T) = n^{-1} \sum_{i=1}^n \xi_{T_i}(t; \boldsymbol{\beta}_T)$ which leads to the Gehan estimating function, and $W_R(\cdot) = W_Y(\cdot) = 1$.

The simulation results are summarized in Table 3.1. Scenario 0 simulated the special case of type (i) longitudinal measurements with the constant samping rate overtime. Note that $R^*(\cdot)$ was not influenced by $\boldsymbol{Z}$ in this case and we let $\boldsymbol{\beta}_R$ equal $\boldsymbol{\beta}_T$ for the reason mentioned in the Section 3.2.1. The samping process considered in Scenario 1 followed a stationary Poisson process and was independent after conditioning on $\boldsymbol{Z}$. In Scenarios 2-3, T were correlated with both $R^*(\cdot)$ and $Y(\cdot)$. Scenario 1 and 2 involved a continuous and a discrete one-dimensional covariate separately. Scenario 3 mimicked a two-arm clinical trial with one covariate from Bernoulli distribution and another covariate from uniform distribution. We used the empirical bias, standard errors and 95% bootstrap confidence intervals for $(\boldsymbol{\beta}_T, \boldsymbol{\beta}_R, \boldsymbol{\beta}_Y)$ to evaluate the performance of the estimation method. The bootstrap procedure was based on $1,000$ replications, where subjuects were sampled with replacement and parameters $\{\boldsymbol{\beta}_T, \boldsymbol{\beta}_R, \boldsymbol{\beta}_Y\}$ were estimated for each replication. As shown in Table 3.1, the estimating method performed well in all the situations considered here. In particular, the 95% bootstrap confidence interval of $\boldsymbol{\beta}_Y$ has empirical coverage probabilities ranging from 0.925 to 0.954 which accords with the asymptotic normality property.

Our method can handle the joint modeling of recurrent events and longitudinal measurements data in the absence of informative terminal event by letting $\boldsymbol{\beta}_T = \mathbf{0}$. We also conducted simulations on sensitivity analysis under the Scenario 2 to illustrate what would happen if the dependence with failure times was ignored. Particularly, we ignored the first layer AFT model and set $\beta_T = 0$ in the second and third layers of the Model (3.2). The simulation results for β_R and β_Y were summarized in Table 3.2, which shows that ignoring dependence of data with failture times can lead significantly biased results if the dependence exists.

Table 3.1: Summary Statistics of the Simulation Studies

n	$\boldsymbol{\beta}_T$ *Bias*	$\boldsymbol{\beta}_T$ *SE*	$\boldsymbol{\beta}_T$ *CP*	$\boldsymbol{\beta}_R$ *Bias*	$\boldsymbol{\beta}_R$ *SE*	$\boldsymbol{\beta}_R$ *CP*	$\boldsymbol{\beta}_Y$ *Bias*	$\boldsymbol{\beta}_Y$ *SE*	$\boldsymbol{\beta}_Y$ *CP*
Scenario 0:	$Z \sim B(1, \frac{1}{2})$, $V \sim N(0, \frac{1}{4})$, $\tau = 7$, $h(t, V) = 0.015$, $a(t) = t^2 + t$, $\mu(V) = V$, $\sigma^2(V) = V^2$, $\log C \sim N[Z, (\lvert Z\rvert + 1)^2]$								
	$\beta_T = -0.5$			$\beta_R = -0.50$			$\beta_Y = -1$		
$n = 100$	−0.004	0.121	0.944	−0.008	0.252	0.939	0.013	0.442	0.940
$n = 200$	−0.000	0.083	0.962	−0.010	0.171	0.944	0.027	0.317	0.952
	$\beta_T = 1$			$\beta_R = 1$			$\beta_Y = 2$		
$n = 100$	0.006	0.130	0.949	0.026	0.211	0.942	−0.001	0.367	0.941
$n = 200$	0.003	0.091	0.942	0.009	0.146	0.937	0.004	0.262	0.944
Scenario 1:	$Z \sim U[0, 1]$, $V \sim Exp(1)$, $\tau = 7$, $h(t, V) = 0.01$, $a(t) = \sin(t)$, $\mu(V) = V$, $\sigma^2(V) = V^2$, $C \sim U[0, 12]$ if $Z < 0.5$ and $C \sim U[0, 15]$ if $Z \geq 0.5$								
	$\beta_T = 1$			$\beta_R = 0$			$\beta_Y = 0$		
$n = 100$	0.002	0.220	0.944	0.028	0.351	0.937	0.028	0.680	0.937
$n = 200$	0.002	0.144	0.962	0.010	0.237	0.946	0.008	0.472	0.941
	$\beta_T = 0.5$			$\beta_R = -0.5$			$\beta_Y = 3$		
$n = 100$	0.003	0.226	0.941	0.005	0.352	0.941	−0.009	0.715	0.925
$n = 200$	−0.002	0.147	0.949	−0.005	0.253	0.943	−0.002	0.502	0.935
Scenario 2:	$Z \sim B(1, \frac{1}{2})$, $V \sim N(0, \frac{1}{4})$, $\tau = 7$, $h(t, V) = \sqrt{3t\lvert V\rvert}/50$, $a(t) = t^2$, $\mu(V) = \sqrt{\lvert V\rvert}$, $\sigma^2(V) = \lvert V\rvert \wedge 1$, $C \sim U[0, 12]$ if $Z = 1$ and $C \sim U[0, 15]$ if $Z = 0$								
	$\beta_T = 1$			$\beta_R = -1$			$\beta_Y = 2$		
$n = 100$	0.002	0.111	0.937	−0.000	0.335	0.941	−0.021	0.470	0.938
$n = 200$	−0.006	0.077	0.943	−0.008	0.224	0.951	−0.028	0.334	0.952
	$\beta_T = -0.5$			$\beta_R = 0$			$\beta_Y = -1$		
$n = 100$	−0.003	0.107	0.944	−0.018	0.271	0.957	0.005	0.464	0.942
$n = 200$	0.003	0.072	0.959	−0.007	0.189	0.955	0.043	0.350	0.942
Scenario 3:	$Z_1 \sim B(1, \frac{1}{2})$, $Z_2 \sim U[0, 1]$, $V \sim N(0, \frac{1}{4})$, $\tau = 7$, $h(t, V) = \frac{\lvert tV\rvert}{20}$, $a(t) = t$, $\mu(V) = V$, $\sigma^2(V) = V^2$, $C \sim U[0, 12]$ if $Z_1 = 0$ and $C \sim U[0, 15]$ if $Z_1 = 1$								
	$\boldsymbol{\beta}_T = (1, 0.5)'$			$\boldsymbol{\beta}_R = (-0.8, -0.5)'$			$\boldsymbol{\beta}_Y = (2, 1)'$		
$n = 100$	0.002	0.111	0.938	−0.005	0.467	0.940	−0.032	0.446	0.936
	−0.007	0.193	0.948	0.014	0.744	0.957	−0.036	0.757	0.950
$n = 200$	0.005	0.076	0.952	0.030	0.320	0.941	−0.021	0.331	0.928
	0.005	0.134	0.953	0.028	0.523	0.958	−0.020	0.550	0.942
	$\boldsymbol{\beta}_T = (-0.3, 0)'$			$\boldsymbol{\beta}_R = (-1, -0.1)'$			$\boldsymbol{\beta}_Y = (-3, 0)'$		
$n = 100$	−0.002	0.108	0.948	−0.014	0.416	0.950	−0.037	0.444	0.937
	0.009	0.181	0.959	0.028	0.709	0.949	0.004	0.711	0.954
$n = 200$	0.001	0.071	0.962	0.006	0.272	0.966	−0.016	0.305	0.942
	0.008	0.131	0.949	0.017	0.493	0.945	−0.005	0.517	0.946

Note: *Bias*, the empirical bias; *SE*, the empirical standard error; *CP*, the empirical coverage probability of 95% bootstrap confidence interval; $B(1, \frac{1}{2})$, the Bernoulli distribution; $N(\mu, \sigma^2)$, the normal distribution; $U[l_1, l_2]$, the Uniform distribution; $Exp(\lambda)$, the Exponential distribution.

Table 3.2: Summary Statistics of the Sensitivity Study under Scenario 2

	$\boldsymbol{\beta}_R$			$\boldsymbol{\beta}_Y$		
n	*Bias*	*SE*	*CP*	*Bias*	*SE*	*CP*
	$\{\beta_T, \beta_R, \beta_Y\} = \{1, -1, 2\}$					
$n = 100$	-1.823	0.318	0.001	-2.788	1.178	0.021
$n = 200$	-1.821	0.227	0.000	-3.169	1.109	0.000
	$\{\beta_T, \beta_R, \beta_Y\} = \{-0.5, 0, -1\}$					
$n = 100$	0.904	0.192	0.016	1.482	0.582	0.023
$n = 200$	0.902	0.135	0.000	1.609	0.479	0.000

Note: *Bias*, the empirical bias; *SE*, the empirical standard error; *CP*, the empirical coverage probability of 95% bootstrap confidence interval.

3.6 Data Analysis

3.6.1 BIOCARD Data Analysis

As an example of type (i) longitudinal measurements, we consider the application of our models on the data from the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) cohort study which aims at identifying biomarkers associated with the development of Alzheimer's Disease (AD) procession. The study was administrated by NIH from 1995 to 2005, and was re-established by a research group at Johns Hopkins School of Medicine after being stopped for four years. Subjects enrolled in the study were cognitively normal at baseline and data including cognitive performance testing scores were collected annually or per half year during the study.

The $\varepsilon 4$ allele of the apolipoprotein E (ApoE) gene is the main genetic risk factor associated with AD dementia (Farrer et al., 1997[6]). Our main object here is to estimate the effect of the ApoE4 gene on the time onset of clinical symptoms and the cognitive performance testing score *Logical Memory IIA - delayed*. Here we only considered the data collected at the even follow-up years in 1995-2005. Totally, we had 236 subjects, consisting of 71 ApoE4 carriers and 165 ApoE4 non-carriers, and the overall censoring rate is 73.7%. Figure 3.6.1 shows the BIOCARD data of uncensored subjects in forward and backward time scales. We first analyzed the data as the right-censored case where the individual enrollment time was the time origin. Let T_i be the

time from entry to the onset of symptoms for mild cognitive impairment (MCI), $R_i(\cdot)$ be the sampling process since entry and $Y_i(\cdot)$ be the cognitive score, $i = 1, \ldots, 236$. This longitudinal measurement is selected because it is found to be a highly predictive marker for onset of symptoms. Since subjects were enrolled at different baseline ages, the centered baseline age of each individual was considered as covariate. By defining the failure time T_i as the age at onset of symptoms, the second set of data analysis is conducted by treating the observed data as being left-truncated and right-censored, where the truncation time L_i is an individual's age at the time when she or he entered the study. In this analysis ApoE4 status was the only covariate.

The analysis results are reported in Table 3.3. For the right-censored data setting, the ApoE4 gene type has a time rescaling effect identified by the AFT model. The positive ApoE4 gene type accelerates the progress of disease and, as a result, the time to onset of symptoms is shortened by the rate $\exp(-0.209)$ for subjects with the same baseline age. Furthermore, the score in the ApoE4 carriers population at time $t \exp(-0.209)$ averagely equals to that in the ApoE4 non-carriers population at time t with an additional increase by a coefficient with value 0.560. Of note, β_Y represents the addition shift effect after being adjusted by the time-rescaling effect, and therefore these two effects should be considered together to understand the covariate effect. Neither the time rescaling effect nor the additional shift effect of ApoE4 gene type showed significant effects according to the 95% bootstrap confidence intervals with $1,000$ bootstrap repetitions. In contrast, the increase of baseline age significantly accelerates the disease process and reduces the score as shown in Table 3.3.

For the left-truncated and right-censored data setting, using a subject's age at onset of symptoms as the failure time, the positive ApoE4 gene type accelerates the progress to disease and shortens the age of onset of symptoms by rate $\exp(-0.044)$. Furthermore, the ApoE4 gene type influenced the rescaled recurrent event proess with an additional multiplier term $\exp(-0.035)$. As to the effect of ApoE4 on the score,

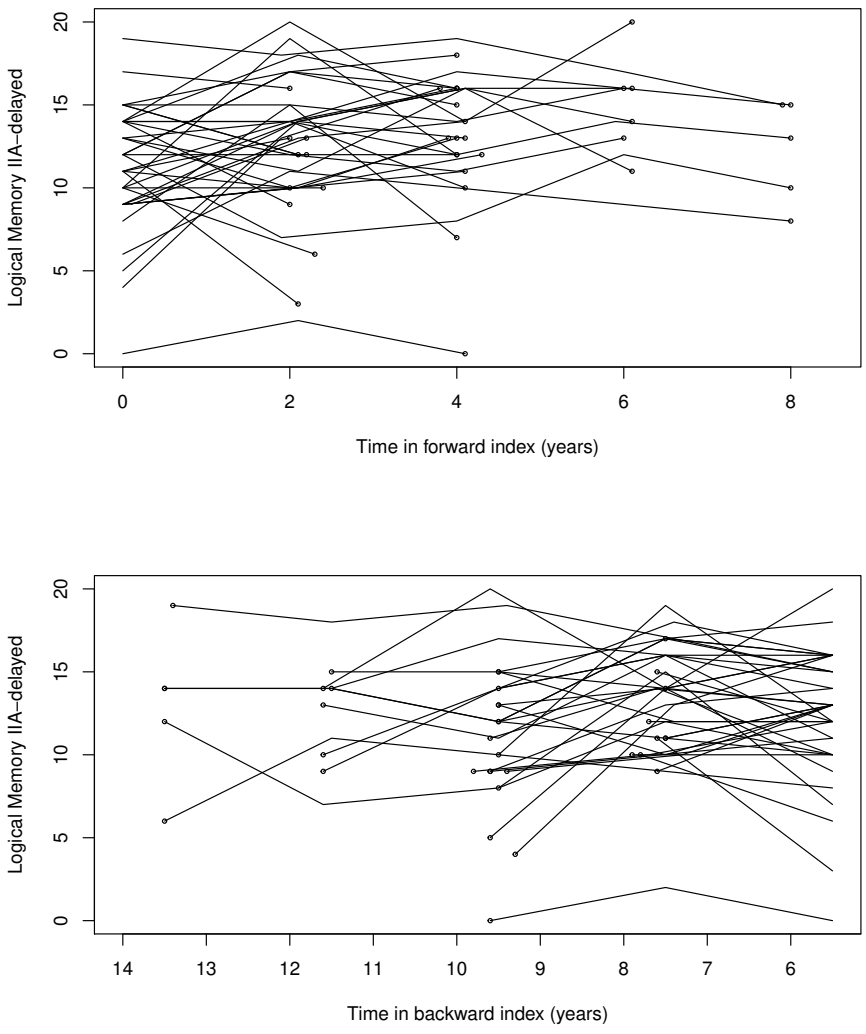

Figure 3.2: Illustration of BIOCARD Data for Uncensored Subjects in Forward and Backward Time Indices

Table 3.3: Summary of BIOCARD Data Analysis

		Coefficient	95% CI
Case 1	Right-Censoring		
$\boldsymbol{\beta}_T$	ApoE4+	-0.209	$(-0.558, 0.167)$
	Baseline Age	-0.040	$(-0.065, -0.022)$
$\boldsymbol{\beta}_R$	ApoE4+	-0.121	$(-0.543, 0.321)$
	Baseline Age	-0.029	$(-0.065, -0.005)$
$\boldsymbol{\beta}_Y$	ApoE4+	0.560	$(-2.001, 3.166)$
	Baseline Age	0.161	$(-0.028, 0.547)$
Case 2	Left Truncation and Right-Censoring		
β_T	ApoE4+	-0.044	$(-0.125, 0.029)$
β_R	ApoE4+	-0.035	$(-0.250, 0.236)$
β_Y	ApoE4+	0.831	$(-2.998, 2.515)$

Note: 95% CI, bootstrap percentile 95% confidence interval.

it provides an additional shift effect by 0.831, but nonsignificantly.

3.6.2 CPCRA Data Analysis

To illustrate the application of the proposed method on type (ii) longitudinal measurement, we analyzed data from a clinical trial conducted by Terry Beirn Community Programs for Clinical Research on AIDS. The study compared two different treatments, didanosine (ddI) and zalcitabine (ddC), for HIV-infected patients who had previously failed treatment with or were intolerant to zidovudine (Abrams et al., 1994[1]). In the trial, 230 patients were randomly sellected to receive ddI treatment and 237 to receive ddC, the event of death is the primary endpoint, and patients were followed until death or censoring. During the course of the trial 363 opportunistic infections were observed, of which 172 were in the ddI group and 191 in the ddc group. For each opportunistic infection, a severity score was provided by physicians as an indicator for the disease progression (Neaton et al., 1994[28]). The analysis in Abrams et al. (1994)[1] suggested that ddC may have provided a survival advantage over ddI.

Table 3.4: Summary of CPCRA Data Analysis

	Coefficient	95% CI
β_T	0.179	$(0.012, 0.328)$
γ	0.156	$(-0.025, 0.407)$
β_Y	0.268	$(-0.733, 0.831)$

Note: 95% CI, bootstrap percentile 95% confidence interval.

We investigated the effect of treatment where the indicator, Z, was the only covariate in the model (coded as 0 for ddI and 1 for ddC). Let T, $R^*(\cdot)$ and $Y(\cdot)$ respectively be the time to death, the opportunistic infection process and the severity score. The analysis results are summarized in Table 3.4, which shows that the treatment ddC had a significant time rescaling effect and elongated the survival time by the rate $\exp(0.179)$. The infection rate of the ddC treatment group at time $t\exp(0.179)$ is equal to that of the ddI treatment group at time t with an additional multiplier $\exp(0.156)$, which is not significant. Moreover, the mean severity score of the ddC treatment group at time $t\exp(0.179)$ is equal to that of the ddI treatment group at time t with an additional increase 0.268 which is also not significant.

3.7 Discussion

In this chapter we developed a joint model for longitudinal measurements, recurrent events, and failure events data where all of the three components of data are treated as outcomes. Without requiring restrict assumptions on recurrent event processes, the proposed model is applicable to both types of longitudinal measurements, (i) and (ii). On the basis of the survivors' population, the model avoids the disputing assumption on the existence of recurrent events or longitudinal measurements after the failure event. As the model does not involve latent variables, computationally it is simpler and easier to adopt when comparing to the shared frailty model. Moreover, the proposed model and estimation inference can be generalized to analyze left-truncated

and right-censored data.

The proposed model involves semiparametric structure in each of the three sub-models, for failure time, recurrent events and longitudinal measurements, and the baseline functions in the model are unspecified. Since the main interest of this chapter is to model and estimate covariates effects, we will only briefly describe the approaches for estimating the baseline functions in Section 3.8.4.

Our model possesses a specific feature that the forward time model is equivalent to the backward-in-time model for recurrent events and longitudinal measurements, where the two models share the same regression parameter values. Therefore, our model can be used to study the terminal behavior of biological processes, such as the performance of a biomarker measurement before the diagnosis of disease or the medicine cost distributions before death. Of note, studying the terminal behavior of a longitudinal measurement process is challenging as the censoring cannot be handled using the standard approach (Chan and Wang, 2016). The proposed model possesses equivalence between forward and backward time scales, which is an attractive feature when modeling stochastic processes (such as recurrent events, longitudinal or functional measurements) in the presence of a terminal event. It has the obvious advantage that one can build a forward time model which is also valid in backward time scale.

Our model also has some limitations. Firstly, our model does not consider the case of time-dependent covariates. Secondly, our model depends on the time rescaling assumption and especially it assumes that the time rescaling rate is uniform for the longitudinal, recurrent event and failure time data. Thirdly, our model needs the independent censoring assumption which is usually not satisfied in follow-up studies and competing risk occurs. For example, subjects are more likely to drop out of studies as their cognition falls sufficiently low. It is not porper to use our model if these assumptions are violated, but unfortunately we have not developed the model

diagnosis or assumption testing methods. Efforts should be made to solve these problems in the future.

3.8 Proofs

3.8.1 Proof of Equivalence of Model (3.2) and (3.3)

First, we prove that Model (3.3) can be induced from Model (3.2). We have

$$\begin{aligned}
&\mathrm{E}\left\{R_0^B(dt^B;T_0)|T_0\right\}\\
=&\mathrm{E}\left\{R_0^*(T_0)-R_0^*((T_0-t^B)^-)|T_0\right\}-\mathrm{E}\left\{R_0^*(T_0)-R_0^*((T_0-t^{B-})^-)|T_0\right\}\\
=&\mathrm{E}\left\{R_0^*(T_0-t^B)-R_0^*((T_0-t^B)^-)|T_0\right\}\\
=&\mathrm{E}\left\{R_0^*(d(T_0-t^B))|T_0\right\}.
\end{aligned}$$

Similarly, we can prove

$$\mathrm{E}\left\{R^B\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}}dt^B;T\right)\mid T,\boldsymbol{Z}\right\}=\mathrm{E}\left\{R^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}}d(T_0-t^B)\right)\mid T,\boldsymbol{Z}\right\},$$

and therefore, by Model (3.2), $E\left\{R^B\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}}dt^B;T\right)\mid T,\boldsymbol{Z}\right\}=e^{\boldsymbol{\beta}_R'\boldsymbol{Z}}\mathrm{E}\left\{R_0^B\left(dt^B;T_0\right)\mid T_0\right\}$. A similar argument extends to $Y^B(\cdot)$ with the same result.

By the symmetric argument, we can prove that Model (3.2) can be induced from Model (3) as well.

3.8.2 Proof of Equation (3.7)

We firstly prove Equation (3.7) under Model (1). If $\xi_{Ri}(t;\boldsymbol{\beta}_T)=0$, the equation (3.7) holds since $dN_{Ri}(t;\boldsymbol{\beta}_T)=0$. If $\xi_{Ri}(t;\boldsymbol{\beta}_T)=1$, we have

$$
\begin{aligned}
&\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}dN_{Ri}(t;\boldsymbol{\beta}_T)\mid \xi_{Ri}(t;\boldsymbol{\beta}_T)=1,\boldsymbol{Z}_i\right]\\
=&\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}\left\{R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}t\right)-R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}t^-\right)\right\}\mid T_i\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\right.\\
&\left.C_{Ri}\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\boldsymbol{Z}_i\right]\\
=&\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}\left\{R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}t\right)-R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}t^-\right)\right\}\mid T_i\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\boldsymbol{Z}_i\right]\\
=&\mathrm{E}\left[\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}dR_i^*(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt)\mid dR_i^*(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt),T_i\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\boldsymbol{Z}_i\right]\mid\right.\\
&\left.T_i\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\boldsymbol{Z}_i\right]\\
=&\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}\mid dR_i^*(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt)=1,T_i\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\boldsymbol{Z}_i\right]\\
&\mathrm{E}\left\{R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt\right)\mid T_i\geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\boldsymbol{Z}_i\right\}\\
=&\alpha_0(t)\cdot e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_i}\mathrm{E}\left\{dR_{0i}^*(t)\mid T_{0i}\geq t\right\}\\
=&\alpha_0(t)\cdot e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_i}\frac{\mathrm{E}\left\{dR_{0i}^*(t)\right\}}{\mathbb{P}\left\{T_{0i}\geq t\right\}}\\
=&d\mathscr{A}_0(t).
\end{aligned}
$$

Thus, we have proved the desired results which implies that $M_{Yi}(t;\mathscr{A}_0,\boldsymbol{\beta}_Y,\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ is a mean-zero stochastic process.

The Equation (3.7) also holds under Model (3.2) which is conditioning on T. We

let $d\mathscr{A}_0(t) = \mathrm{E}\left[Y_0(t)R_0^*(dt) \mid T_0 \geq t\right]$ and have

$$
\begin{aligned}
&\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right) - \boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\} dN_{Ri}(t;\boldsymbol{\beta}_T) \mid \xi_{Ri}(t;\boldsymbol{\beta}_T) = 1, \boldsymbol{Z}_i\right] \\
=&\mathrm{E}\left[\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right) - \boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\}\left\{R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}t\right) - R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}t^-\right)\right\} \mid T_i \geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i},\right. \\
&\left. C_{Ri} \geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}, \boldsymbol{Z}_i\right] \\
=&\mathrm{E}\left[\mathbb{1}_{\left[T_i \geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right]}\mathrm{E}\left\{\left(Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right) - \boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right)R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt\right) \mid T_i, \boldsymbol{Z}_i\right\}\right] \\
&\left\{\mathbb{P}\left(T_i \geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i} \mid \boldsymbol{Z}_i\right)\right\}^{-1} \\
=&\mathrm{E}\left[\mathbb{1}_{\left[T_i \geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right]}\mathrm{E}\left\{\left(Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right) - \boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right) \mid R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt\right) = 1, T_i, \boldsymbol{Z}_i\right\}\right. \\
&\left.\mathrm{E}\left\{R_i^*\left(e^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}dt\right) \mid T_i, \boldsymbol{Z}_i\right\}\right]\left\{\mathbb{P}\left(T_i \geq te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i} \mid \boldsymbol{Z}_i\right)\right\}^{-1} \\
=&\mathrm{E}\left[\mathbb{1}_{[T_{0i} \geq t]}\mathrm{E}\left\{Y_{0i}\left(t\right) \mid T_{0i}, R_{0i}^*\left(dt\right) = 1\right\}\mathrm{E}\left\{R_{0i}^*\left(dt\right) \mid T_{0i}\right\}\right]\left\{\mathbb{P}\left(T_{0i} \geq t\right)\right\}^{-1} \\
=&\mathrm{E}\left[Y_{0i}(t)R_{0i}^*(dt) \mid T_{0i} \geq t\right] \\
=&d\mathscr{A}_0(t).
\end{aligned}
$$

3.8.3 Proof of Theorem 3.1

First we mainly prove that $n^{-1/2}\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ converges weakly to a tight zero mean Gaussian variable. Simple algebraic manipulation yields that

$$
\begin{aligned}
n^{-1/2}\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R) = &\int_0^\infty dn^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0) \\
&-\sum_{i=1}^n\int_0^\infty \tilde{\boldsymbol{Z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)dn^{-1/2}\sum_{i=1}^n M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0).
\end{aligned}
$$

Since both $\sum_{i=1}^n M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0)$ and $\sum_{i=1}^n \boldsymbol{Z}_i M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0)$ are sums of n i.i.d. zero-mean terms, $n^{-1/2}\sum_{i=1}^n M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0)$ and $n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i M_{Yi}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R,\boldsymbol{\beta}_Y,\mathscr{A}_0)$ converge weakly to tight zero mean Gaussian processes with continuous sample paths by conditions (C1-2). Since $\boldsymbol{Z}_i$ is assumed to be bounded in condition (C2), one can prove that $\tilde{\boldsymbol{Z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ converges in probability to a deterministic function, $\tilde{\boldsymbol{z}}(t;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ say, uniformly in t. By repeatedly

using Lemma 1 in Lin et al. (2000)[21], one can prove that $n^{-1/2}\boldsymbol{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)$ converges weakly to a tight zero mean Gaussian variable. Similar arguments extend to the convergence of $n^{-1/2}\boldsymbol{U}_T(\boldsymbol{\beta}_T)$ (Ying, 1993[49]) and $n^{-1/2}\boldsymbol{U}_R(\boldsymbol{\beta}_R;\boldsymbol{\beta}_T)$ (Lin et al., 2000[21]; Huang and Wang, 2003[11]) based on conditions (C1-4). So $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta})$ converge weakly to a multivariate normal random variable with mean zero and variance denoted as Σ (Van Der Vaart and Wellner, 1996[39]; Lin et al., 2000[21]).

Define

$$\begin{aligned}
\mathscr{U}_T(\boldsymbol{\beta}_T) &= \sum_{i=1}^{n}\int_0^{\infty}\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\mathrm{E}\left\{\xi_{Tj}(t;\boldsymbol{\beta}_T)\right\}\boldsymbol{Z}_j}{\sum_{j=1}^{n}\mathrm{E}\left\{\xi_{Tj}(t;\boldsymbol{\beta}_T)\right\}}\right] d\mathrm{E}\left\{N_{Ti}(t;\boldsymbol{\beta}_T)\right\},\\
\mathscr{U}_R(\boldsymbol{\beta}_R;\boldsymbol{\beta}_T) &= \sum_{i=1}^{n}\int_0^{\infty}\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\mathrm{E}\left\{\xi_{Rj}(t;\boldsymbol{\beta}_T)\right\}e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}{\sum_{j=1}^{n}\mathrm{E}\left\{\xi_{Rj}(t;\boldsymbol{\beta}_T)\right\}e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}\right] d\mathrm{E}\left\{N_{Ri}(t;\boldsymbol{\beta}_T)\right\},\\
\mathscr{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R) &= \sum_{i=1}^{n}\int_0^{\infty}\left[\boldsymbol{Z}_i - \frac{\sum_{j=1}^{n}\boldsymbol{Z}_j\mathrm{E}\left\{\xi_{Rj}(t;\boldsymbol{\beta}_T)\right\}e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}{\sum_{j=1}^{n}\mathrm{E}\left\{\xi_{Rj}(t;\boldsymbol{\beta}_T)\right\}e^{\boldsymbol{\beta}_R'\boldsymbol{Z}_j}}\right]\left\{Y_i\left(te^{\boldsymbol{\beta}_T'\boldsymbol{Z}_i}\right)\right.\\
&\quad \left.-\boldsymbol{\beta}_Y'\boldsymbol{Z}_i\right\} d\mathrm{E}\left\{N_{Ri}(t;\boldsymbol{\beta}_T)\right\}.
\end{aligned}$$

Let $\mathscr{U}(\boldsymbol{\beta}) = \left(\mathscr{U}_T(\boldsymbol{\beta}_T)', \mathscr{U}_R(\boldsymbol{\beta}_R;\boldsymbol{\beta}_T)', \mathscr{U}_Y(\boldsymbol{\beta}_Y;\boldsymbol{\beta}_T,\boldsymbol{\beta}_R)'\right)'$. By the conditions (C3-4), $\mathrm{E}\left\{\xi_{Ti}(t;\boldsymbol{\beta}_T)\right\}$, $\mathrm{E}\left\{\xi_{Ri}(t;\boldsymbol{\beta}_T)\right\}$, $\mathrm{E}\left\{N_{Ti}(t;\boldsymbol{\beta}_T)\right\}$ and $\mathrm{E}\left\{N_{Ri}(t;\boldsymbol{\beta}_T)\right\}$ as functions of $\boldsymbol{\beta}_T$ satisfy continuity and derivative properties. According to the conditions (C1-5) and applying the techniques of Ying (1993)[49], one can prove that there exists a matrix $\boldsymbol{A}_n$ such that for every sequence $d_n > 0$ with $d_n \to 0$ in probability, we have

$$\sup_{\|\boldsymbol{b}-\boldsymbol{\beta}\|\le d_n} \| \boldsymbol{U}(\boldsymbol{b}) - \boldsymbol{U}(\boldsymbol{\beta}) - \boldsymbol{A}_n n(\boldsymbol{b}-\boldsymbol{\beta}) \| / (\sqrt{n} + n \| \boldsymbol{b}-\boldsymbol{\beta} \|) = o_p(1).$$

If the eigenvalues of $\boldsymbol{A}_n$ are all bounded away from zero for all large enough n and $\boldsymbol{A}_n \to A$ where A is nonsingular, there exists a closed neighborhood $\mathcal{N}$ containing $\boldsymbol{\beta}$ as its interior point such that $\boldsymbol{\beta}$ is strongly consistent by Ying (1993)[49]. Additionally,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -n^{-1/2}\boldsymbol{A}_n^{-1}\boldsymbol{U}(\boldsymbol{\beta}) + o_p(1) \to N\left(\boldsymbol{0}, \boldsymbol{A}^{-1}\Sigma(\boldsymbol{A}^{-1})'\right)$$

weakly in $\mathcal{N}$.

3.8.4 The Inference of Baseline Functions

For Model (3.1) and (3.2), with estimated regression parameters, we can estimate $\mathscr{A}_0(t)$ by (3.10), and estimate $\Lambda_{R_0}(t)$ by a Breslow-type estimator (Breslow, 1972)

$$\hat{\Lambda}_{R_0}(t;\hat{\boldsymbol{\beta}}_T) = \sum_{i=1}^{n} \int_0^t \frac{dN_{Ri}\left(u;\hat{\boldsymbol{\beta}}_T\right)}{\sum_{j=1}^{n} \xi_{Rj}\left(u;\hat{\boldsymbol{\beta}}_T\right) e^{\hat{\boldsymbol{\beta}}_R^{'}\boldsymbol{Z}_j}}.$$

For Model (3.1), $\alpha_0(t) = \mathrm{E}\left\{Y_0(t)|T_0 \geq t, R_0^*(dt) = 1\right\}$ can be estimated by Kernel Smoothing method (Ramlau-Hansen, 1983)

$$\hat{\alpha}_0(t;\hat{\boldsymbol{\beta}}_T) = \frac{\int_0^{+\infty} K_h(t-s) d\hat{\mathscr{A}_0}(s;\hat{\boldsymbol{\beta}}_T)}{\int_0^{+\infty} K_h(t-s) d\hat{\Lambda}_{R_0}(s;\hat{\boldsymbol{\beta}}_T)},$$

where $K_h(x) = h^{-1}K(x/h)$ is a kernel function with bandwidth h, $\int_{-1}^{1} K(x)dx = 1$, and $\int_{-1}^{1} xK(x)dx = 0$. Using the empirical process theory (Van Der Vaart and Wellner, 1996[39]; Kosorok, 2008[15]) and Kernel Smoothing techniques (Ramlau-Hansen, 1983[30]) one can show that each of the processes $n^{1/2}\{\hat{\mathscr{A}_0}(t;\hat{\boldsymbol{\beta}}_T) - \mathscr{A}_0(t)\}$, $n^{1/2}\{\hat{\Lambda}_{R_0}(t;\hat{\boldsymbol{\beta}}_T) - \Lambda_{R_0}(t)\}$ and $(nh)^{1/2}\{\hat{\alpha}_0(t;\hat{\boldsymbol{\beta}}_T) - \alpha_0(t)\}$ converges weakly to mean-zero Gaussian processes by weak convergence theory for functional parameters. Since these baseline functions are not the focus of our work, we skip details of the asymptotic proofs.

To illustrate the performance of the estimation of $\alpha_0(\cdot)$ in Model (1), we conducted the simulation studies under the Scenario 0 (see Table 3.1) where data were generated exactly followed the Model (1) and

$$\begin{aligned}\alpha_0(t) =& \frac{\mathrm{E}\left[\mathbb{1}_{[T_0 \geq t]} \mathrm{E}\left\{Y_0(t)|T_0, R_0^*(dt) = 1\right\}\right]}{\mathbb{P}(T_0 \geq t|R_0^*(dt) = 1)} \\ =& a(t) + \frac{1}{2\pi}\exp\left\{-2\left(\log(t)\right)^2\right\}\left\{1 - \Phi(2\log(t))\right\}^{-1}.\end{aligned}$$

Here $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. As presented in the above equation, part of the variation of $\alpha_0(\cdot)$ (i.e. the residual of $\alpha_0(t) - a(t)$) was from the change of the survivor's population with different times, and this happens because of the correlation between $Y_0(\cdot)$ and T_0. Here, we used

Table 3.5: Summary Statistics of the Simulation Studies of Estimating $\alpha_0(\cdot)$ in Model (1) under Scenario 0

	$t = 0.5$		$t = 1$		$t = 1.5$	
n	$Bias$	SE	$Bias$	SE	$Bias$	SE
			$\{\beta_T, \beta_R, \beta_Y\} = \{-0.5, -0.5, -1\}$			
$n = 100$	0.013	0.201	-0.003	0.284	-0.0480	0.660
$n = 200$	0.004	0.144	-0.002	0.204	0.009	0.457
			$\{\beta_T, \beta_R, \beta_Y\} = \{1, 1, 2\}$			
$n = 100$	0.029	0.201	0.056	0.295	0.067	0.670
$n = 200$	0.014	0.147	0.019	0.194	0.035	0.435

Note: $Bias$, the empirical bias; SE, the empirical standard error.

the Epanechnikov Kernel (i.e. $K(x) = 3/4(1 - x^2)\mathbb{1}_{[|x|\leq 1]}$) with the bandwidth $h = n^{-1/3}$. We summarized the simulation results in the Table 3.5 and confirmed good performance of the estimation results.

Chapter 4

Change Point Estimation in Backward Process Model

4.1 INTRODUCTION

In studies of disease progression, it is often of interest to study biomarker performance prior to the occurrence of failure events by aligning failure events as time origins and counting time backward. For example, in studies of Alzheimer's Disease the rate of change in biomarker measurement before diagnosis of disease is widely recognized as an important index for predicting the disease (Hall et al., 2000[8]; Wilson et al., 2007[43]). However, the conventional forward perspective of stochastic processes are not designed for the terminal behavior of processes.

Additionally, when studying trajectories of biomarkers or other longitudinal measurements before the failure event such as death, researchers sometimes find the phenomenon that the rate of change of biomarkers would start to shift, accelerate or decelerate, at some special time point which is prior to the failure event by a gap time. We name this special type of change point as backward change point to distinguish it from the traditional forward change point which occurs after the origin point by a constant period of time. Backward change point is of scientific interest for biology progression study and disease diagnosis. In particular, for Alzheimer's disease (AD) studies, Jack et al. (2010)[12] hypothesized that the decline of biomarkers began to

accelerate prior to diagnosis in order such that different biomarkers characterized the disease during different stages. Knowing the acceleration order of biomarkers decline will help to measure disease progression precisely, and then therapeutic intervention can be given to patients properly.

Consider the model where the failure time is T and the decline of biomarker begins to accelerate at time point $T - d$ for those subjects with failure time longer than d, where d is a constant gap time. Ideally, if all of the subjects are followed until diagnosis, the backward change point model can be analyzed by aligning the biomarker measurements retrospectively with diagnosis as the time origin, and then use techniques for forward change point models (e.g., Slate and Cronin, 1997[32]; Skates, Pauler, and Jacobs ,2001[31]) to estimate the parameter d.

However, in most follow-up studies, subjects may be censored due to design limitations, early drop out or other reasons. For censored subjects, not only the time of failure event is missing, the longitudinal measurements between censoring and failure event are also unobserved. As a result, the new time origins of censored data cannot be set and the observed longitudinal measurements cannot be indexed in backward time index. Thus, the censoring problem of longitudinal and survival analysis studies in backword time index is more complex than that of the studies in forward time index. However, as a characteristic of time-to-event data, disregarding the censoring problem will lead to biased results even if the cenoring is noninformative. To our knowledge, there are few valid methods of modeling backward change point problems. Some researchers (Hall et al., 2000[8]; Wilson et al., 2011[44]) have already found the scientific meaning of backward change point in applications, but did not consider the censoring issue in their models. Additionally, longitudinal measurements and the failure event are usually correlated and the association within data structure should also be considered. Unfortunately, the current methods of joint modeling of lingitudinal measuremetns and time-to-event data (Wulfsohn and Tsiatis, 1997[45];

Henderson et al., 2000[9]; Xu and Zeger, 2001[47]; Song et al., 2002[33]; Vonesh et al., 2006[40]; Song and Wang, 2008[34]) do not consider the backward change point problem.

In this chapter, we develop a method to model the backward change point problem. Our method is inspired by the backward process estimation methods of Chan and Wang (2010)[4] where stochastic counting processes are modeled by aligning failure events as time origins. The general idea of our method is that we model the longitudinal measurement process retrospectively starting from the failure event and the failure time will be a condition of the model just like a covariate. We present our model in Section 4.2. The estimation method and asymptotic properties are shown in Section 4.3. We conducted simulation studies to explore the finite sample property of our estimation method which is presented in Section 4.4. The data analysis is presented in Section 4.5 as an illustration. Some extra discussions are shown in Section 4.6.

4.2 BACKWARD CHANGE POINT MODEL

To set notation, denote by $\boldsymbol{Z}$ a $p \times 1$ vector of covariates, T the time to the failure event, and $R(t)$ the number of sampling times at or before time t where t is the forward time index. The longitudinal process $Y(t)$ is measured at sampling times where $R(dt) = 1$. Let C be the censoring time for the failure event and longitudinal measurements. When framing the model, we only consider longitudinal measurements before the failure time, i.e. $0 \leq t \leq T$. For backward process, denote by t^B the backward time index. Define the backward process of longitudinal measurements as $Y^B(t^B; T) = Y(T - t^B)$ and the backward sampling process as $R^B(t^B; T) = R(T) - R((T - t^B)^-)$, where $0 \leq t^B$. In practice, it is possible to consider $t^B > T$. For example, in Alzheimer's Disease, some individuals may experience the shifts of biomarkers before entry into the study, but the backward change point as

the main interest can be estimated by our model in population level if some other individuals' longitudinal measurements before backward change point are available. We assume that the backward change point is localized in the study follow-up time interval in this chapter.

Consider the following assumptions:

(A1) $Y(\cdot)$ is independent of C conditioning on $\{R(dt) = 1, T, \boldsymbol{Z}\}$.

Assumption (A1) is proposed in forward time model, which is convenient in terms of assumption understanding and interpretation because the data is generated in forward time index. Assumption (A1) implies that $Y^B(\cdot)$ is independent of C conditioning on $\left\{R^B(dt^B; T) = 1, T, \boldsymbol{Z}\right\}$.

In diseases progressions, biomarker variables usually change gradually along continous trajectories over time. Under Assumption (A1), we first consider the backward change point model for this situation as

$$\mathrm{E}\left\{Y^B(t^B; T)|R^B(dt^B; T) = 1, T, \boldsymbol{Z}\right\} = \psi_0(t^B, T, \boldsymbol{Z}; \boldsymbol{\beta}) + \eta(t^B - d)\mathbb{1}_{[t^B > d]}, \quad (4.1)$$

where $t^B \geq 0$ and $\psi_0(t^B, T, \boldsymbol{Z}; \boldsymbol{\beta})$ is a prespecified regression function with parameter $\boldsymbol{\beta}$. The backward change point parameter is d and the corresponding shift rate parameter is η. We assume that $\eta \neq 0$ to avoid the inidentifiability issue. It is natural to develop the model conditioning on T, like a covariate, since the whole backward processes are defined based on T. The condition event $R^B(dt^B; T) = 1$ states that Model (1) models the mean trajectory and covariate effects only for the observed $Y^B(\cdot; T)$. This is a natual condition for marker measurements observed when recurent events occur. For continuous longitudinal measurements the result generalizes to the underling $Y^B(\cdot)$ if the longitudinal measurements and recurrent event processes are independent with each other given $\boldsymbol{Z}$ and T. The model can be interpreted in backward time index as that for those subjects with failure time longer than d the slope of $Y^B(t^B; T)$ will have an additional change by η when $t^B > d$.

This is equivalent to that the slope of $Y(t)$ will have an additional change by $-\eta$ in $[T-d, T]$ compared to that in $[0, T-d)$ in forward time index. To simplify notations, we define $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta, d)'$ and $\psi(t^B, T, \boldsymbol{Z}; \boldsymbol{\theta}) = \psi_0(t^B, T, \boldsymbol{Z}; \boldsymbol{\beta}) + \eta(t^B - d)\mathbb{1}_{[t^B > d]}$.

As an instance, we can simply define a linear regression function, termed as Model (A),

$$\psi_A(t^B, T, \boldsymbol{Z}; \boldsymbol{\theta}) = \beta_0 + \beta_1 T + \beta_2 t^B + \boldsymbol{\beta}_3^T \boldsymbol{Z} + \eta(t^B - d)\mathbb{1}_{[t^B > d]},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \boldsymbol{\beta}_3^T)$. In backward time index, this model has the failure event as the time origin and the change point is located at $t^B = d$ for subjects with $T > d$. The failure time T has a linear effect on the intercept value, i.e. $Y^B(0; T)$. The slope of $Y^B(t^B; T)$ is β_2 between the failure event and the change point (i.e. $0 \leq t^B \leq d$) and $(\beta_2 + \eta)$ after the change point (i.e. $t^B > d$). When in forward time index, the change point is at $(T - d)$ and the slope of $Y(\cdot)$ is $-(\beta_2 + \eta)$ for $t < T - d$ and $-\beta_2$ for $t \in [T-d, T]$. Besides this linear regression function, other regression functions of $\psi_0(t^B, T, \boldsymbol{Z}; \boldsymbol{\beta})$ can be adopted as long as it has proper interpretations and satisfies the analysis requirement. Generally, we can define a more flexible model by letting $\boldsymbol{\theta}$ be a prespecified function of $\{T, \boldsymbol{Z}\}$, i.e. $\boldsymbol{\theta}(T, \boldsymbol{Z})$, where different subjects are allowed to have different slopes and change points based on their $\{T, \boldsymbol{Z}\}$.

In some situations, researchers also have interest in longitudinal processes with sudden-jump change points and this type of trajectory has a discrete change point. Under Assumption (A1), we can also consider another model as:

$$\mathrm{E}\left\{Y^B(t^B; T) | R^B(dt^B; T) = 1, T, \boldsymbol{Z}\right\} = \psi_0(t^B, T, \boldsymbol{Z}; \boldsymbol{\beta}) + \eta_1 \mathbb{1}_{[t^B > d]} + \eta_2 (t^B - d)\mathbb{1}_{[t^B > d]}, \tag{4.2}$$

where there is an additional jump η_1 at the change point and the shift rate parameter is η_2. We assume that $\eta_1 \neq 0$ for modeling identifiability issue. To simplify notations, we let $\boldsymbol{\eta} = (\eta_1, \eta_2)'$ and let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\eta}', d)'$ stand for all of the parameters in Model (4.2). Without specification, $\boldsymbol{\eta}$ can also stand for η in Model (4.1). Similarly, a simple

linear model example can be

$$\psi_A(t^B, T, \boldsymbol{Z}; \boldsymbol{\theta}) = \beta_0 + \beta_1 T + \beta_2 t^B + \boldsymbol{\beta}_3^T \boldsymbol{Z} + \eta_1 \mathbb{1}_{[t^B > d]} + \eta_2 (t^B - d) \mathbb{1}_{[t^B > d]}$$

which is denoted as Model (B).

Our Models (4.1) and (4.2) allow the correlation between C and T even though conditioning on $\boldsymbol{Z}$. This good property will help to avoid complex debate when considering competing risk problems where failure events caused by other reasons can happen before our target failure event and therefore lead to censoring. For example, subjects may die due to heart disease or cancer before having Alzheimer's disease. This kind of censoring time C is usually dependent of the target failure time T and therefore the independent censoring assumption will not be accepted when competing risk problem occurs. Because the independent censoring is not required, our Models (4.1) and (4.2) can provide unbiased analysis even if competing risk problem exists.

4.3 ESTIMATION

4.3.1 Estimation Methods

Model (4.1) and (4.2) share the same estimating procedure which will be introduced in this section. Define $\tilde{T} = T \wedge C$ and $\Delta = \mathbb{1}_{[T \leq C]}$. The observations of $\{R(t), Y(t)\}$ will stop at $t = \tilde{T}$. We define the counting process of failure event by $N(t) = \mathbb{1}_{\left[\tilde{T} \leq t\right]} \Delta$, and the risk set indicator function by $\xi(t) = \mathbb{1}_{\left[\tilde{T} \geq t\right]}$. Suppose the observations $\boldsymbol{X}_i = \{\tilde{T}_i, \Delta_i, R_i(\cdot), Y_i(\cdot), Z_i\}$, $i = 1, \ldots, n$, are independent and identically distributed. We assume that d belongs to a compact set $[0, \tau]$ and $\mathbb{P}(\tilde{T} > d) > 0$.

For any $0 \leq t$ and $0 \leq t^B$, we define

$$M_i(t, t^B; \boldsymbol{Z}_i, \boldsymbol{\theta}) = \int_0^t \int_0^{t^B} \mathbb{1}_{[u \leq s]} \left\{ Y_i^B(u; s) - \psi(u, s, \boldsymbol{Z}_i; \boldsymbol{\theta}) \right\} R_i^B(du; s) N_i(ds). \quad (4.3)$$

One can prove that $\mathrm{E}\left\{ M_i(dt, dt^B; \boldsymbol{Z}_i, \boldsymbol{\theta}) | T_i, \boldsymbol{Z}_i \right\} = 0$ (see the Appendix). We adopt the traditional least square method to estimate the parameters in Model (4.1 - 4.2).

In particular, we define

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\left\{Y_i^B(t^B;t) - \psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta})\right\}^2 R_i^B(dt^B;t)N_i(dt), \tag{4.4}$$

and then we find the estimator of $\boldsymbol{\theta}$ such that it minimizes $Q_n(\boldsymbol{\theta})$, i.e. $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} Q_n(\boldsymbol{\theta})$. Denote $\boldsymbol{\gamma} = (\boldsymbol{\beta}^T, \boldsymbol{\eta}^T)^T$. One can first fix d and calculate the profile minimum point $\hat{\boldsymbol{\gamma}}(d) = \arg\min_{\boldsymbol{\gamma}} Q_n(\boldsymbol{\gamma}, d)$, and then find $\hat{d}$ such that $\hat{d} = \arg\min_d Q_n(\hat{\boldsymbol{\gamma}}(d), d)$ and $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}(\hat{d})^T, \hat{\boldsymbol{\eta}}(\hat{d})^T, \hat{d}\right)^T$.

As an illustration, we adopt the above method to estimate parameters in the linear regression change point Model (A). Denote $\boldsymbol{L}_i(t^B,t;\boldsymbol{Z}_i,d) = \left(1, t, t^B, \boldsymbol{Z}_i^T, (t^B-d)\mathbb{1}_{[t^B>d]}\right)^T$. We first fix $d \in [0,\tau]$ and minimizes $Q_n(\boldsymbol{\gamma}, d)$ as a function of $\boldsymbol{\gamma}$. In particular, the partial derivative function is

$$\frac{\partial Q_n(\boldsymbol{\gamma},d)}{\partial \boldsymbol{\gamma}} = \frac{1}{n}\sum_{i=1}^{n}\int_0^{+\infty}\int_0^{\infty} \boldsymbol{L}_i(t^B,t;\boldsymbol{Z}_i,d)M_i(dt,dt^B;\boldsymbol{Z}_i,\boldsymbol{\theta}). \tag{4.5}$$

Let $\hat{\boldsymbol{\gamma}}(d)$ be the solution of $\partial Q_n(\boldsymbol{\gamma},d)/\partial\boldsymbol{\gamma} = \mathbf{0}$, which has a closed form solution

$$\begin{aligned}\hat{\boldsymbol{\gamma}}(d) = &\left\{\sum_{i=1}^{n}\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\boldsymbol{L}_i(t^B,t;\boldsymbol{Z}_i,d)\boldsymbol{L}_i^T(t^B,t;\boldsymbol{Z}_i,d)R_i^B(dt^B;t)N_i(dt)\right\}^{-1}\\ &\left\{\sum_{i=1}^{n}\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\boldsymbol{L}_i(t^B,t;\boldsymbol{Z}_i,d)Y_i^B(t^B;t)R_i^B(dt^B;t)N_i(dt)\right\},\end{aligned} \tag{4.6}$$

if $\left\{\sum_{i=1}^{n}\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\boldsymbol{L}_i(t^B,t;\boldsymbol{Z}_i,d)\boldsymbol{L}_i^T(t^B,t;\boldsymbol{Z}_i,d)R_i^B(dt^B;t)N_i(dt)\right\}$ is invertible. We then estimate d as

$$\hat{d} = \arg\min_{d\in[0,\tau]} Q_n(\hat{\boldsymbol{\gamma}}(d), d), \tag{4.7}$$

by grid search. Finally, we have the least squre estimator $\hat{\boldsymbol{\theta}} = \left(\hat{\boldsymbol{\beta}}(\hat{d})^T, \hat{\eta}(\hat{d}), \hat{d}\right)^T$.

4.3.2 Asymptotic Properties

We first introduce the inference properties of Model (4.1). We define

$$q(\boldsymbol{\theta};\boldsymbol{X}) = \int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\left(Y^B(t^B;t) - \psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta})\right)^2 R^B(dt^B;t)N(dt),$$

and $\mathcal{Q}(\boldsymbol{\theta}) = \mathrm{E}\{q(\boldsymbol{\theta}; \boldsymbol{X})\}$, where $\mathrm{E}\{\cdot\}$ means the expectation under true parameters $\boldsymbol{\theta}_0$. We define $\rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|$ where $\| \cdot \|$ stands for the L_2 norm. The consistency property and the weak convergence property of $\hat{\boldsymbol{\theta}}$ will be established under the Assumption (A1) and the following conditions:

(i) $\boldsymbol{\theta}_0$ lies in a compact set $\boldsymbol{\Theta}$ in the Euclidean space.

(ii) $\{\boldsymbol{Z}, T, R(\cdot), Y\}$ is bounded and $\{\boldsymbol{Z}, T\}$ has a positive and continuous density function.

(iii) $\psi_0(t^B, T, \boldsymbol{Z}; \boldsymbol{\beta})$, is uniformly bounded, has continuous second order differential function with respect to $\boldsymbol{\beta}$, belongs to Donsker class, and $\psi(t^B, T, \boldsymbol{Z}; \boldsymbol{\theta})$ is identifiable.

(iv) $\eta \neq 0$ in Model (4.1).

(v) Define

$$\boldsymbol{A}_1(t^B, t, \boldsymbol{\theta}) = \left(\frac{\partial \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}, 0, 0 \right)^T$$

and

$$\boldsymbol{A}_2(t^B, t, \boldsymbol{\theta}) = \left(\frac{\partial \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}, (t^B - d), -\eta \right)^T$$

. There exists a neighborhood of $\boldsymbol{\theta}_0$ such that

$$\mathrm{E}\left\{ \int_0^{+\infty} \int_0^{+\infty} \boldsymbol{A}_1(t^B, t, \boldsymbol{\theta}) \boldsymbol{A}_1(t^B, t, \boldsymbol{\theta})^T \mathbb{1}_{[t^B \leq d_0]} + \boldsymbol{A}_2(t^B, t, \boldsymbol{\theta}) \boldsymbol{A}_2(t^B, t, \boldsymbol{\theta})^T \mathbb{1}_{[t^B > d_0]} \right.$$
$$\left. R^B(dt^B; t) N(dt) \right\}$$

is positive definite.

(vi) For all n large enough and sufficiently small δ,

$$\mathrm{E}^* \sup_{\rho_1(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \delta} \sqrt{n} |(Q_n - \mathcal{Q})(\boldsymbol{\theta}) - (Q_n - \mathcal{Q})(\boldsymbol{\theta}_0)| \leq K\delta,$$

where K is a finite positive constant.

The conditions (i-ii) are commonly-used regular conditions for random variables of stochastic processes. The conditions (iii-vi) primary mean that the model function $\psi_0(\cdot)$ should be relatively smooth, the derived derivative function matrix is non-sigular, and the corresponding $Q_n(\cdot)$ is easy to handle. We present the consistency of $\hat{\boldsymbol{\theta}}$ in Theorem (4.3.1). Theorem (4.3.2) states that the rates of convegence of $\hat{\boldsymbol{\theta}}$ is $\sqrt{n}$, and Theorem (4.3.3) proves that $\hat{\boldsymbol{\theta}}$ has normal limiting distributions.

Theorem 4.3.1 *Under conditions (i-iii), there exists a neighborhood $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ such that $\hat{\boldsymbol{\theta}}$ lies in $\boldsymbol{\Theta}$, and then it converges in probability to $\boldsymbol{\theta}_0$ as $n \to \infty$.*

Theorem 4.3.2 *Under conditions (i-vi), $\sqrt{n}\rho_1(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_P(1)$.*

Theorem 4.3.3 *Under conditions (i-vi), $\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ converges weakly to a mean zero multivariate Gaussian variable.*

Secondly, we introduce the inference properties of Model (4.2). We define $\rho_2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \sqrt{\| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \|_2^2 + \| \hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0 \|_2^2 + |\hat{d} - d_0|}$. Beside conditions (i-iii), we also need the following conditions:

(iv*) $\eta_{10} \neq 0$ in Model (4.2).

(v*) Define $\boldsymbol{A}_3(t^B, t, \boldsymbol{\theta}) = \left(\frac{\partial \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}, 1, (t^B - d), -\eta_2\right)^T$. There exists a neighborhood of $\boldsymbol{\theta}_0$ such that

$$\mathrm{E}\left\{\int_0^{+\infty}\int_0^{+\infty} \boldsymbol{A}_1(t^B, t, \boldsymbol{\theta})\boldsymbol{A}_1(t^B, t, \boldsymbol{\theta})^T \mathbb{1}_{[t^B \leq d_0]} + \boldsymbol{A}_3(t^B, t, \boldsymbol{\theta})\boldsymbol{A}_3(t^B, t, \boldsymbol{\theta})^T \mathbb{1}_{[t^B > d_0]} \right.$$
$$\left. R^B(dt^B; t) N(dt)\right\}$$

is positive definite.

(vi*) For all n large enough and sufficiently small δ,

$$\mathrm{E}^* \sup_{\rho_2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) < \delta} \sqrt{n}|(Q_n - \mathcal{Q})(\boldsymbol{\theta}) - (Q_n - \mathcal{Q})(\boldsymbol{\theta}_0)| \leq K\delta,$$

where K is a finite positive constant.

(vii*) $\boldsymbol{I}_1(\boldsymbol{\theta}) = \mathrm{E}[-\frac{\partial^2 q(\boldsymbol{\theta};\boldsymbol{X})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}]$ and $\boldsymbol{I}_2(\boldsymbol{\theta}) = \mathrm{Var}[\frac{\partial q(\boldsymbol{\theta};\boldsymbol{X})}{\partial\boldsymbol{\gamma}}]$ exist, and $\boldsymbol{I}_1(\boldsymbol{\theta}_0)$ is non-singular.

We present the consistency of $\hat{\boldsymbol{\theta}}$ in Theorem (4.3.4). Theorem (4.3.5) states that the rates of convegence of $\hat{d}$ and $\hat{\boldsymbol{\gamma}}$ are n and $\sqrt{n}$ separately. Since the convergence rate of $\hat{d}$ is faster than that of $\hat{\boldsymbol{\gamma}}$, the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ is still the norm distribution which is shown in Theorm (4.3.6).

Theorem 4.3.4 *Under conditions (i-iii), there exists a neighborhood $\boldsymbol{\Theta}$ of $\boldsymbol{\theta}_0$ such that $\hat{\boldsymbol{\theta}}$ lies in $\boldsymbol{\Theta}$, and then it converges in probability to $\boldsymbol{\theta}_0$ as $n \to \infty$.*

Theorem 4.3.5 *Under conditions (i-iii, iv*-vii*), $\sqrt{n}\rho_2(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) = O_P(1)$.*

Theorem 4.3.6 *Under conditions (i-iii, iv* - vii*),*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \to \boldsymbol{N}\left(\mathbf{0}, \boldsymbol{I}_1(\theta_0)^{-1}\boldsymbol{I}_2(\theta_0)\boldsymbol{I}_1(\theta_0)^{-1}\right)$$

weakly in $\boldsymbol{\Theta}$.

Achieving the weak convergence property for $\hat{d}$ in Model (4.2) is challenging. The reason is that traditional weak convergence results of M-estimator are based on Taylor expansions of the induced estimating equations (e.g. the score function in the maximum likelihood estimation method) in a small neighborhood of the true value. While, the function $Q_n(\boldsymbol{\theta})$ in our method is not differentiable over the change point parameter d and therefore the traditional technique can not be used. Some researchers (Pons, 2003[29]; Kosorok and Song, 2007[16]) found that the asymptotic normality does not hold for the change point parameter in their problem settings.

4.4 Simulation Studies

We conducted abundant simulation studies to evaluate the performance of the proposed estimating method for both Model (4.1) and (4.2). The Model (4.1) and (4.2)

shared the same data generation procedures. Let τ be the maximum follow-up time. The ith $(i = 1, \ldots, n)$ individual's data were generated by the following steps:

- Generate the covariate vector $\boldsymbol{Z}_i$. Here, we only considered two covariates which had Bernoulli and standard normal distributions separately.

- Generate the potential failure time T_i and censoring time C_i given $\boldsymbol{Z}_i$. Since T_i and C_i were potentially correlated with each other, we generated a random vector $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2})^T$ from a multivariate normal distribution where the mean and covariance matrix were functions of $\boldsymbol{Z}_i$, denoted as $\boldsymbol{\mu}_T(\boldsymbol{Z}_i)$ and $\boldsymbol{\Sigma}_T(\boldsymbol{Z}_i)$ separately. Define $T_i = \exp(\omega_{i1})$ and $C_i = \exp(\omega_{i2})$.

- Generate the sampling time process $R_i(\cdot)$ on $[0, \tau]$ from stationary Poisson process which was independent with $\{\boldsymbol{Z}_i, T_i, C_i\}$.

- Given T_i and $\boldsymbol{Z}_i$, generate the longitudinal variable at each sampling time in backward direction. Here, we only considered linear regression models, i.e. Model (A) and (B) for continuous and discrete types of backward change point models separately. The error term $\varepsilon_i(t)$ had normal distribution $N(0, \sigma^2(\boldsymbol{Z}_i, T_i))$, where $\sigma^2(\cdot, \cdot)$ was a prespecified function.

The simulation procedure follows the Model (4.1) and (4.2) exactly and was determined by prespecified components: sample size n, τ, functions $\{\mu_T(\cdot), \boldsymbol{\Sigma}_T(\cdot), \sigma^2(\cdot, \cdot)\}$, and parameters $(d, \boldsymbol{\beta}', \eta)'$. We simulated multiple type of data sets by using different prespecified components where $1,000$ simulated data sets were repeatedly generated for each case.

The simulation results were summarized in Table 4.1-4.2 for Model (4.1)and Table 4.3-4.4 for Model (4.2). In Scenario 1, the variance of the error term $\varepsilon_i(t)$ depended on both of the failure time T and covariates $\boldsymbol{Z}$, and the failure time were correlated with censoring time C. While in Scenario 2, $Y(\cdot)$ was independent with T conditioning on

Table 4.1: Summary Statistics of the Simulation Studies of Scenario 1 for Model (4.1)

n		$\boldsymbol{\beta}^T$	η	d
Scenario 1:		$\boldsymbol{\mu}(\boldsymbol{Z}) = (Z_1+Z_2, Z_1+Z_2+2)'$, $\sigma^2(\boldsymbol{Z},T) = \min(\frac{1}{5}T\exp(Z_1+Z_2),1)$		
		$\boldsymbol{\Sigma}_{1,1} = \max((\lvert Z_1+Z_2\rvert+1)/2,1)$, $\boldsymbol{\Sigma}_{2,2} = \max(\exp(Z_1+Z_2)/3,1)$,		
		$\boldsymbol{\Sigma}_{1,2} = \boldsymbol{\Sigma}_{2,1} = \sqrt{\lvert\boldsymbol{\Sigma}(\boldsymbol{Z})[1,1]\boldsymbol{\Sigma}(\boldsymbol{Z})[2,2]-0.5\rvert}$, $\tau = 15$.		
		$\boldsymbol{\beta}^T = (4.5, 3, -3, 4, -4.5)$	$\eta = 5$	$d = 4$
	B	$(0.0001, -0.0001, 0.0006, -0.0018, -0.0009)$	-0.0028	-0.0008
$n = 100$	V	$(0.0867, 0.0185, 0.0529, 0.0994, 0.0599)$	0.0676	0.0409
	CR	$(0.930, 0.893, 0.929, 0.911, 0.914)$	0.917	0.934
	B	$(0.0028, -0.0003, 0.0000, -0.0017, -0.0008)$	0.0001	0.0004
$n = 500$	V	$(0.0380, 0.0079, 0.0228, 0.0419, 0.0254)$	0.0273	0.0168
	CR	$(0.955, 0.932, 0.944, 0.937, 0.945)$	0.941	0.961
		$\boldsymbol{\beta}^T = (1, 0.5, 1.5, -0.5, -1.5)$	$\eta = -1$	$d = 2$
	B	$(-0.0028, -0.0002, 0.0077, -0.0022, -0.0010)$	-0.0096	-0.0156
$n = 100$	V	$(0.1071, 0.0186, 0.1177, 0.0999, 0.0603)$	0.1171	0.1956
	CR	$(0.930, 0.894, 0.908, 0.911, 0.913)$	0.925	0.895
	B	$(0.0021, -0.0003, 0.0014, -0.0018, 0.0007)$	-0.0018	0.0026
$n = 500$	V	$(0.0437, 0.0079, 0.0478, 0.0420, 0.0255)$	0.0481	0.0796
	CR	$(0.952, 0.931, 0.944, 0.934, 0.946)$	0.944	0.930

Note: B, the empirical bias; V, the empirical standard error; CR, the coverage rate of 95% confidence interval.

$\boldsymbol{Z}$, and T and C are uncorrelated. We used the empirical bias, standard errors, and 95% confidence interval coverage rate to evaluate the performance of the estimation method. The 95% confidence intercal of $\hat{d}$ in Model (4.2) was calculated by using the simple Bootstrap method. As shown in Table 4.1-4.4, the estimating method performed well in all of the situations considered here. In particular, to compare the convergence rates of $\hat{d}$ in Model (4.1) and (4.2), we calculated the variance of $\sqrt{n}(\hat{d}-d_0)$ in Model (4.1) and $n(\hat{d}-d_0)$ in Model (4.2) and ploted the density functions of $\hat{d}-d_0$ with multiple sample sizes (see Table 4.5 and Figure 4.1). The results were consistent with our statements.

4.5 Data Analysis

In this section we consider the application of our models on the data from the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) cohort study which aims to identify biomarkers associated with the development of Alzheimer's Disease (AD). The study was conducted by NIH from 1995 to 2005, and was re-

Table 4.2: Summary Statistics of the Simulation Studies of Scenario 2 for Model (4.1)

n		$\boldsymbol{\beta}^T$	η	d
Scenario 2:		$\boldsymbol{\mu}(\boldsymbol{Z}) = (Z_1 - Z_2 + 1, Z_1 Z_2 + 4)'$, $\sigma^2(\boldsymbol{Z}, T) = \frac{1}{10}(\lvert Z_1 Z_2 \rvert + 1)$		
		$\boldsymbol{\Sigma}_{1,1} = (\lvert Z_1 - Z_2 \rvert + 1)/2$, $\boldsymbol{\Sigma}_{2,2} = (\lvert Z_1 + Z_2 \rvert + 1)/2$,		
		$\boldsymbol{\Sigma}_{1,2} = \boldsymbol{\Sigma}_{2,1} = 0$, $\tau = 10$.		
		$\boldsymbol{\beta}^T = (4.5, 3, -3, 4, -4.5)$	$\eta = 5$	$d = 4$
	B	$(0.0004, -0.0001, 0.0002, 0.0000, 0.0006)$	0.0008	0.0004
$n = 100$	V	$(0.0379, 0.0081, 0.0156, 0.0306, 0.0227)$	0.0278	0.0133
	CR	$(0.938, 0.917, 0.943, 0.938, 0.919)$	0.923	0.916
	B	$(0.0006, -0.0000, -0.0003, -0.0014, 0.0003)$	-0.0002	-0.0003
$n = 500$	V	$(0.0161, 0.0033, 0.0068, 0.0130, 0.0093)$	0.0111	0.0062
	CR	$(0.951, 0.942, 0.954, 0.944, 0.952)$	0.949	0.892
		$\boldsymbol{\beta}^T = (1, 0.5, 1.5, -0.5, -1.5)$	$\eta = -1$	$d = 2$
	B	$(0.0005, -0.0002, 0.0003, 0.0002, 0.0005)$	-0.0004	0.0015
$n = 100$	V	$(0.0473, 0.0081, 0.0363, 0.0306, 0.0228)$	0.0380	0.0560
	CR	$(0.923, 0.918, 0.944, 0.935, 0.921)$	0.935	0.922
	B	$(0.0001, -0.0000, 0.0004, -0.0001, 0.0003)$	-0.0004	-0.0008
$n = 500$	V	$(0.0197, 0.0033, 0.0158, 0.0130, 0.0093)$	0.0159	0.0247
	CR	$(0.947, 0.941, 0.950, 0.949, 0.955)$	0.954	0.933

Note: B, the empirical bias; V, the empirical standard error; CR, the coverage rate of 95% confidence interval.

Table 4.3: Summary Statistics of the Simulation Studies of Scenario 1 for Model (4.2)

n		$\boldsymbol{\beta}^T$	$\boldsymbol{\eta}^T$	d
Scenario 1:		$\boldsymbol{\mu}(\boldsymbol{Z}) = (Z_1 + Z_2, Z_1 + Z_2 + 2)'$, $\sigma^2(\boldsymbol{Z}, T) = \min(\frac{1}{5} T \exp(Z_1 + Z_2), 1)$		
		$\boldsymbol{\Sigma}_{1,1} = \max((\lvert Z_1 + Z_2 \rvert + 1)/2, 1)$, $\boldsymbol{\Sigma}_{2,2} = \max(\exp(Z_1 + Z_2)/3, 1)$,		
		$\boldsymbol{\Sigma}_{1,2} = \boldsymbol{\Sigma}_{2,1} = \sqrt{\lvert \boldsymbol{\Sigma}(\boldsymbol{Z})[1,1] \boldsymbol{\Sigma}(\boldsymbol{Z})[2,2] - 0.5 \rvert}$, $\tau = 15$.		
		$\boldsymbol{\beta}^T = (4.5, 3, -3, 4, -4.5)$	$\boldsymbol{\eta}^T = (3.5, 5)$	$d = 4$
	B	$(-0.0001, -0.0001, 0.0008, -0.0019, -0.0012)$	$(-0.0020, -0.0032)$	-0.0013
$n = 100$	V	$(0.0864, 0.0186, 0.0527, 0.0993, 0.0599)$	$(0.2553, 0.0679)$	0.0321
	CR	$(0.936, 0.892, 0.926, 0.913, 0.912)$	$(0.862, 0.916)$	1.000
	B	$(0.0028, -0.0003, 0.0001, -0.0017, 0.0008)$	$(-0.0014, 0.0000)$	0.0000
$n = 500$	V	$(0.0382, 0.0079, 0.0228, 0.0419, 0.0255)$	$(0.0889, 0.0273)$	0.0054
	CR	$(0.954, 0.931, 0.948, 0.937, 0.945)$	$(0.938, 0.944)$	0.999
		$\boldsymbol{\beta}^T = (1, 0.5, 1.5, -0.5, -1.5)$	$\boldsymbol{\eta}^T = (-1, -2)$	$d = 2$
	B	$(-0.0059, -0.0001, 0.0123, -0.0016, -0.0012)$	$(-0.0003, -0.0115)$	-0.0083
$n = 100$	V	$(0.1058, 0.0186, 0.1176, 0.1002, 0.0606)$	$(0.2339, 0.1197)$	0.0892
	CR	$(0.936, 0.888, 0.918, 0.912, 0.908)$	$(0.881, 0.902)$	0.978
	B	$(0.0019, -0.0003, 0.0016, -0.0017, 0.0007)$	$(-0.0009, -0.0017)$	-0.0006
$n = 500$	V	$(0.0429, 0.0079, 0.0463, 0.0421, 0.0255)$	$(0.0781, 0.0471)$	0.0154
	CR	$(0.951, 0.931, 0.948, 0.935, 0.948)$	$(0.935, 0.946)$	0.991

Note: B, the empirical bias; V, the empirical standard error; CR, the coverage rate of 95% confidence interval.

Table 4.4: Summary Statistics of the Simulation Studies of Scenario 2 for Model (4.2)

n		$\boldsymbol{\beta}^T$	$\boldsymbol{\eta}^T$	d
Scenario 2:		$\boldsymbol{\mu}(\boldsymbol{Z}) = (Z_1 - Z_2 + 1, Z_1 Z_2 + 4)'$, $\sigma^2(\boldsymbol{Z}, T) = \frac{1}{10}(\lvert Z_1 Z_2 \rvert + 1)$		
		$\boldsymbol{\Sigma}_{1,1} = (\lvert Z_1 - Z_2 \rvert + 1)/2$, $\boldsymbol{\Sigma}_{2,2} = (\lvert Z_1 + Z_2 \rvert + 1)/2$,		
		$\boldsymbol{\Sigma}_{1,2} = \boldsymbol{\Sigma}_{2,1} = 0$, $\tau = 10$.		
		$\boldsymbol{\beta}^T = (4.5, 3, -3, 4, -4.5)$	$\boldsymbol{\eta}^T = (3.5, 5)$	$d = 4$
	B	$(0.0002, -0.0002, 0.0003, -0.0001, 0.0006)$	$(-0.0054, 0.0008)$	-0.0005
$n = 100$	V	$(0.0380, 0.0801, 0.0154, 0.0306, 0.0227)$	$(0.0947, 0.0278)$	0.0139
	CR	$(0.938, 0.916, 0.950, 0.939, 0.918)$	$(0.823, 0.922)$	1.000
	B	$(0.0005, -0.0000, -0.0003, 0.0014, 0.0003)$	$(0.0012, -0.0002)$	0.0000
$n = 500$	V	$(0.0161, 0.0033, 0.0066, 0.0130, 0.0093)$	$(0.0288, 0.0110)$	0.0018
	CR	$(0.946, 0.943, 0.953, 0.944, 0.953)$	$(0.927, 0.950)$	1.000
		$\boldsymbol{\beta}^T = (1, 0.5, 1.5, -0.5, -1.5)$	$\boldsymbol{\eta}^T = (-1, -2)$	$d = 2$
	B	$(0.0011, -0.0002, -0.0007, 0.0001, 0.0005)$	$(-0.0001, 0.0007)$	0.0005
$n = 100$	V	$(0.0471, 0.0081, 0.0360, 0.0306, 0.0229)$	$(0.0564, 0.0379)$	0.0101
	CR	$(0.923, 0.915, 0.937, 0.936, 0.920)$	$(0.929, 0.933)$	1.000
	B	$(0.0001, -0.0000, 0.0003, -0.0001, 0.0003)$	$(-0.0012, -0.0003)$	0.0000
$n = 500$	V	$(0.0196, 0.0033, 0.0156, 0.0130, 0.0093)$	$(0.0244, 0.0158)$	0.0013
	CR	$(0.947, 0.940, 0.954, 0.949, 0.955)$	$(0.936, 0.955)$	1.000

Note: B, the empirical bias; V, the empirical standard error; CR, the coverage rate of 95% confidence interval.

Table 4.5: Compare the Convergence Rates of $\hat{d}$ in Model (4.1) and (4.2)

n	$\text{Var}\{\sqrt{n}(\hat{d} - d_0)\}$ for Model (4.1)	$\text{Var}\{n(\hat{d} - d_0)\}$ for Model (4.2)
200	0.3205	33.4586
1000	0.2909	30.9200
5000	0.2843	30.5511
10000	0.2823	32.5249

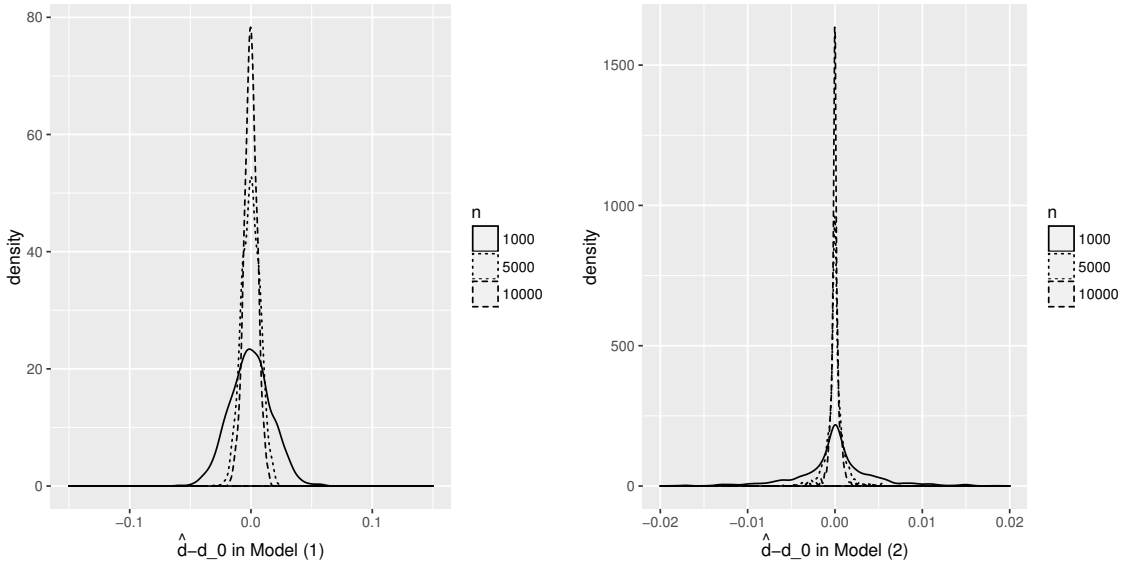

Figure 4.1: Compare the Convergence Rates of $\hat{d}$ in Model (4.1) and (4.2).

Table 4.6: Descriptive Statistics of BIOCARD Data Analysis

#Subjects	#Censored/ #Uncensored	#ApoE4+/ #ApoE4-	#Male/ #Female	Mean Years of Education (SD)	Mean Baseline Age (SD)
289	226/63	91/198	118/171	17.078 (2.371)	56.940 (9.960)

administrated by a research team at Johns Hopkins School of Medicine after since 2009. Subjects enrolled in the study were cognitively normal at entry time and data including cognitinve scores were longitudinally collected during the study.

Our main object is to estimate the backward change point of the cognitive score WAIS-RDigitSymbol before the onset of clinical symptoms of AD by using our model (A). In this analysis, we adjusted the covariate effects of years of education, baseline age, gender, and the $\varepsilon 4$ allele of the apolipoprotein E (ApoE) gene type (encoded as 1 for carriers and 0 for non-carriers) which is the main genetic risk factor associated with AD dementia (Farrer et al., 1997[5]). Totally, we had 289 subjects, consisting of 63 uncensored subjects and 226 censored subjects. More descriptive statistics are presented in Table 4.6.

The analysis results are reported in Table 4.7. In backward time index and with the failure event as the time origin, the change point is located at $\hat{d} = 5.857$. Because the uncensored sample size was small and the measurement of each subject were not dense, the 95% confidence interval of $\hat{d}$ was relative wide comparing to the study follow-up time (20 years). The failure time T together with other covariates has a linear effect on the intercept, i.e. the WAIS-RDigitSymbol score at the failure time. With t^B increasing by one unit, $Y^B(t^B; T)$ increased by 0.082 significantly between the failure event and change point ($0 \leq t^B \leq 5.857$) and increased by 0.005 after the change point ($t^B > 5.857$). However the change shift was not significant according to the 95% confidence interval of $\hat{\eta}$, which is consistent with the wide confidence interval of $\hat{d}$. As the model interpretation in the forward time index, the change point is at

Table 4.7: Summary of BIOCARD Data Analysis

	Estimator	Standard Error	95% Confidence Interval
Intercept	-0.545	0.561	(-1.649, 0.559)
T	-0.012	0.0182	(-0.048, 0.023)
t^B	0.082	0.037	(0.009, 0.155)
Gender (Male)	-0.180	0.142	(-0.458, 0.099)
Baseline age	-0.023	0.009	(-0.041, -0.006)
Education	0.070	0.031	(0.008, 0.132)
ApoE4	0.013	0.150	(-0.282, 0.307)
$\hat{\eta}$	-0.077	0.053	(-0.182, 0.028)
$\hat{d}$	5.857	2.228	(1.470, 10.244)

$T - 5.857$, and the WAIS-RDigitSymbol score decreased by -0.005 before the change point and -0.082 after the change point. The results were mainly consistent with the biological interpretation.

4.6 Discussion

In this chapter, we have proposed a statistical method to study the terminal behavior of stochastic processes and estiamte the backward change point prior to the failure event. Our approach considers the censoring problem which is, to our knowledge, usually ignored when estimating the backward change point. Especially, our method allows the censoring time to be associated with the failure time and therefore can be applied with the occurence of competing risks issue. With the spirit of semiparametric models, our method has few restrictions on the variable distributions or the correlations within the data structure except for a few common assumptions. We also studied the asymptotic properties of our method and confirmed them by simulation studies. We have illustrated our method by applying it to the BIOCARD data.

Unfortunately, our methods take few advantage of the available data of the censored subjects. In many follow up studies, the censoring rate can be very high (e.g. the censoring rate is about 0.8 in Biocard cohort study) and how to use the data from

the censored subjects can be an interesting research topic in the future. Moreover, there have been many studies in developing forward-in-time models for stochastic processes and some researchers have started to consider models in backward time scale, but few efforts have been conducted to combine models in forward and backward together. We think that modeling in both forward and backward time index may be a solution to how to use the data of censored subjects and should be considered as the future work.

4.7 Proofs

4.7.1 Proof Estimation Method

Suppose Model (4.1) is true and $\boldsymbol{\theta}_0$ is the true parameter. If $t^B \leq t$, by the property of conditinal exprectation, we have

$$\begin{aligned} &\mathrm{E}\left\{M_i(dt, dt^B; \boldsymbol{Z}_i, \boldsymbol{\theta}_0)|\boldsymbol{Z}_i\right\} \\ =&\mathrm{E}\left[\mathrm{E}\left\{\left(Y_i^B(t^B;t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right) R_i^B(dt^B;t) N_i(dt) \mid N_i(dt), \boldsymbol{Z}_i\right\} \mid \boldsymbol{Z}_i\right]. \end{aligned} \tag{4.8}$$

If the subject is censored (i.e. $N_i(dt) = 0$), the backward processes can not be identified and therefore the conditionaly expectation is 0. So

$$\begin{aligned} &\mathrm{E}\left[\mathrm{E}\left\{\left(Y_i^B(t^B;t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right) R_i^B(dt^B;t) N_i(dt) \mid N_i(dt), \boldsymbol{Z}_i\right\} \mid \boldsymbol{Z}_i\right] \\ =&\mathrm{E}\left\{\left(Y_i^B(t^B;t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right) R_i^B(dt^B;t) \mid N_i(dt) = 1, \boldsymbol{Z}_i\right\} \mathrm{E}\left\{N_i(dt) \mid \boldsymbol{Z}_i\right\}. \end{aligned} \tag{4.9}$$

Similarly, by applying the property of conditional expectation with conditioning on $R_i^B(dt^B;t)$, we have

$$\begin{aligned} &\mathrm{E}\left[\left(Y_i^B(t^B;t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right) R_i^B(dt^B;t) \mid N_i(dt) = 1, \boldsymbol{Z}_i\right\} \\ =&\mathrm{E}\left[\mathrm{E}\left\{\left(Y_i^B(t^B;t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right) R_i^B(dt^B;t) \mid R_i^B(dt^B;t), N_i(dt) = 1,\right.\right. \\ &\left.\left.\boldsymbol{Z}_i\right\} \mid N_i(dt) = 1, \boldsymbol{Z}_i\right] \\ =&\mathrm{E}\left\{\left(Y_i^B(t^B;t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right) \mid R_i^B(dt^B;t) = 1, N_i(dt) = 1, \boldsymbol{Z}_i\right\} \\ &\mathrm{E}\left\{R_i^B(dt^B;t) \mid N_i(dt) = 1, \boldsymbol{Z}_i\right\} \end{aligned} \tag{4.10}$$

Because of the independent censoring assumption (A1), we have

$$\begin{aligned}
&\mathrm{E}\left\{\left(Y_i^B(t^B;t)-\psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right)\mid R_i^B(dt^B;t)=1,N_i(dt)=1,\boldsymbol{Z}_i\right\}\\
=&\mathrm{E}\left\{\left(Y_i^B(t^B;t)-\psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right)\mid R_i^B(dt^B;t)=1,T_i=t,C_i\geq t,\boldsymbol{Z}_i\right\}\\
=&\mathrm{E}\left\{\left(Y_i^B(t^B;t)-\psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right)\mid R_i^B(dt^B;t)=1,T_i=t,\boldsymbol{Z}_i\right\}.
\end{aligned}\tag{4.11}$$

Based on the model, finally we have

$$\begin{aligned}
&\mathrm{E}\left\{M_i(dt,dt^B;\boldsymbol{Z}_i,\boldsymbol{\theta}_0)|\boldsymbol{Z}_i\right\}\\
=&\mathrm{E}\left\{\left(Y_i^B(t^B;t)-\psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right)\mid R_i^B(dt^B;t)=1,T_i=t,\boldsymbol{Z}_i\right\}\\
&\mathrm{E}\left\{R_i^B(dt^B;t)\mid N_i(dt)=1,\boldsymbol{Z}_i\right\}\mathrm{E}\left\{N_i(dt)\mid\boldsymbol{Z}_i\right\}\\
=&0.
\end{aligned}\tag{4.12}$$

If $t^B>t$, one can easily prove that $\mathrm{E}\left\{M_i(dt,dt^B;\boldsymbol{Z}_i,\boldsymbol{\theta}_0)|\boldsymbol{Z}_i\right\}=0$, and therefore the equation exists as well.

4.7.2 Proof of Theorem (4.3.1) (4.3.4)

Theorem (4.3.1) (4.3.4) share the same proof. By the property of conditional expectation, one can prove that

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta})=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty}\mathbb{1}_{[t^B\leq t]}\mathrm{E}\left\{\left(Y^B(t^B;t)-\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta})\right)^2|T=t,\boldsymbol{Z}\right\}\right.\\
&\left.\mathrm{E}\left\{R^B(dt^B;t)|N(dt)=1,\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right].
\end{aligned}\tag{4.13}$$

Therefore,

$$\begin{aligned}
&\mathcal{Q}(\boldsymbol{\theta})-\mathcal{Q}(\boldsymbol{\theta}_0)\\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty}\mathbb{1}_{[t^B\leq t]}\left(-2\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta})\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta}_0)+\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta})^2+\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta}_0)^2\right)\right.\\
&\left.\mathrm{E}\left\{R^B(dt^B;t)|N(dt)=1,\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right]\\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty}\mathbb{1}_{[t^B\leq t]}\left(\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta})-\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta}_0)\right)^2\mathrm{E}\left\{R^B(dt^B;t)|N(dt)=1,\boldsymbol{Z}\right\}\right.\\
&\left.\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right]\\
\geq&0
\end{aligned}\tag{4.14}$$

where the equation holds if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ based on the identifiable property in condition (iii). Based on conditions (i-iii), $\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta})$ belongs to Donsker class,

Lipschitz continuous functions of Donsker classes are Donsker, and furthermore if a function is Donsker, it is also Glivenko-Cantelli, we know that $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |Q_n(\boldsymbol{\theta}) - \mathcal{Q}(\boldsymbol{\theta})|$ converges in probability to zero as $n \to 0$. Since $\mathcal{Q}(\boldsymbol{\theta})$ is continuous, we have $\hat{\boldsymbol{\theta}}$ converges to $\boldsymbol{\theta}_0$ in probability as $n \to 0$ according to the Argmax Theorem (Kosorok, Theorem 14.1, 2008[17]).

4.7.3 Proof of Theorem (4.3.2)

We have

$$
\begin{aligned}
&\mathcal{Q}(\boldsymbol{\theta}) - \mathcal{Q}(\boldsymbol{\theta}_0) \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\left(\psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta}) - \psi(t^B,t,\boldsymbol{Z};\boldsymbol{\theta}_0)\right)^2 \mathrm{E}\left\{R^B(dt^B;t)|N(dt),\boldsymbol{Z}\right\}\right. \\
&\left.\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\left(\psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}) - \psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}_0)\right)^2 \mathbb{1}_{[t^B\le d\wedge d_0]}\right. \\
&+ \mathbb{1}_{[t^B\le t]}\left(\psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}) + \eta(t^B-d) - \psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}_0)\right)^2 \mathbb{1}_{[d<t^B\le d_0]} \\
&+ \mathbb{1}_{[t^B\le t]}\left(\psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}) - \psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}_0) - \eta_0(t^B-d_0)\right)^2 \mathbb{1}_{[d_0<t^B\le d]} \\
&+ \mathbb{1}_{[t^B\le t]}\left(\psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}) + \eta(t^B-d) - \psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}_0) - \eta_0(t^B-d_0)\right)^2 \mathbb{1}_{[t^B>d\vee d_0]} \\
&\left.\mathrm{E}\left\{R^B(dt^B;t)|N(dt)=1,\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
\ge&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\left(\psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}) - \psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}_0)\right)^2 \mathbb{1}_{[t^B\le d\wedge d_0]}\right. \\
&+ \mathbb{1}_{[t^B\le t]}\left(\psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}) + \eta(t^B-d) - \psi_0(t^B,t,\boldsymbol{Z};\boldsymbol{\beta}_0) - \eta_0(t^B-d_0)\right)^2 \mathbb{1}_{[t^B>d\vee d_0]} \\
&\left.\mathrm{E}\left\{R^B(dt^B;t)|N(dt)=1,\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\left((\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\boldsymbol{A}_1(t^B,t,\boldsymbol{\theta}) + O(\parallel\boldsymbol{\theta}-\boldsymbol{\theta}_0\parallel_2^2)\right)^2 \mathbb{1}_{[t^B\le d\wedge d_0]}\right. \\
&+ \mathbb{1}_{[t^B\le t]}\left((\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\boldsymbol{A}_2(t^B,t,\boldsymbol{\theta}) + O(\parallel\boldsymbol{\theta}-\boldsymbol{\theta}_0\parallel_2^2)\right)^2 \mathbb{1}_{[t^B>d\vee d_0]} \\
&\left.\mathrm{E}\left\{R^B(dt^B;t)|N(dt)=1,\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\left\{\boldsymbol{A}_1(t^B,t,\boldsymbol{\theta})\boldsymbol{A}_1(t^B,t,\boldsymbol{\theta})^T\mathbb{1}_{[t^B\le d\wedge d_0]} + \right.\right. \\
&\left.\left.\boldsymbol{A}_2(t^B,t,\boldsymbol{\theta})\boldsymbol{A}_2(t^B,t,\boldsymbol{\theta})^T\mathbb{1}_{[t^B>d\vee d_0]}\right\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\,\mathrm{E}\left\{R^B(dt^B;t)|N(dt),\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
&+ o(\parallel\boldsymbol{\theta}-\boldsymbol{\theta}_0\parallel_2^2) \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T\left\{\boldsymbol{A}_1(t^B,t,\boldsymbol{\theta})\boldsymbol{A}_1(t^B,t,\boldsymbol{\theta})^T\mathbb{1}_{[t^B\le d_0]} + \right.\right. \\
&\left.\left.\boldsymbol{A}_2(t^B,t,\boldsymbol{\theta})\boldsymbol{A}_2(t^B,t,\boldsymbol{\theta})^T\mathbb{1}_{[t^B>d_0]}\right\}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\,\mathrm{E}\left\{R^B(dt^B;t)|N(dt),\boldsymbol{Z}\right\}\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
&+ o(\parallel\boldsymbol{\theta}-\boldsymbol{\theta}_0\parallel_2^2).
\end{aligned}
$$

Based on the conditions (iv-v), there exists a positive constants c_1 such that

$$\mathcal{Q}(\boldsymbol{\theta}) - \mathcal{Q}(\boldsymbol{\theta}_0) \ge c_1 \parallel\boldsymbol{\theta}-\boldsymbol{\theta}_0\parallel_2^2 \tag{4.15}$$

in a small neighbourhood of $\boldsymbol{\theta}_0$. Because of the condition (vi), the desired convergence rate follows (Kosorok, Theorem 14.4, 2008[17]). Note of, the proofs in Kosorok (2008)[17] was originally for suddern-jump change point models, but we have shown that the techniques can be applied to continuous change point models like Model (4.1)

4.7.4 Proof of Theorem (4.3.3)

Denote $\mathcal{T}_1^B = \{t^B : t^B \leq d \wedge d_0\}$, $\mathcal{T}_2^B = \{t^B : d < t^B \leq d_0\}$, $\mathcal{T}_3^B = \{t^B : d_o < t^B \leq d\}$, and $\mathcal{T}_4^B = \{t^B : t^B > d \vee d_0\}$. We consider $Q_n(\boldsymbol{\theta})$ on the four spaces separately and denote $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_k^B) = \{Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}_0)\} \mathbb{1}_{\left[t^B \in \mathcal{T}_k^B\right]}$ $(k = 1, 2, 3, 4)$. It can be easily shown that

$$
\begin{aligned}
& G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_k^B) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \left[\left\{ Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) \right\}^2 - \left\{ Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right\}^2 \right] \\
& \mathbb{1}_{\left[t^B \in \mathcal{T}_k^B\right]} R_i^B(dt^B; t) N_i(dt) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \left\{ \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta})^2 - 2Y_i^B(t^B; t)\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)^2 + \right. \\
& \left. 2Y_i^B(t^B; t)\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right\} \mathbb{1}_{\left[t^B \in \mathcal{T}_k^B\right]} R_i^B(dt^B; t) N_i(dt) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \left[\left\{ \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right\}^2 \right. \\
& + 2\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta})\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) - 2Y_i^B(t^B; t)\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) \\
& \left. -2\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0)^2 + 2Y_i^B(t^B; t)\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right] \mathbb{1}_{\left[t^B \in \mathcal{T}_k^B\right]} R_i^B(dt^B; t) N_i(dt) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \left[\left\{ \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right\}^2 - \right. \\
& \left. 2 \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \left(\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \right] \\
& \mathbb{1}_{\left[t^B \in \mathcal{T}_k^B\right]} R_i^B(dt^B; t) N_i(dt).
\end{aligned}
\tag{4.16}
$$

We firstly consider $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)$ and divide it into two parts, i.e. $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(1)}$

and $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(2)}$, with

$$\begin{aligned} G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(1)} =& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left\{ \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right\}^2 \\ & \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt), \end{aligned}$$

$$\begin{aligned} G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(2)} = & -\frac{2}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \\ & \left(\psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt). \end{aligned}$$

We will show that $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(1)} = o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2)$ and $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(2)} = o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| / \sqrt{n})$.

Define

$$\begin{aligned} f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) =& \frac{1}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left\{ \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right\}^2 \\ & \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt). \end{aligned}$$

By applying Taylor Expansion on $\psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta})$, there exsits a positive constant c_2 such that

$$\begin{aligned} & f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) \\ =& \frac{1}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left\{ \psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta}) + \eta \left(t^B - d \right) - \psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta}_0) \right\}^2 \\ & \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt) \\ \le & \frac{1}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left\{ \psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta}) - \psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta}_0) \right\}^2 + \\ & \eta^2 \left(t^B - d \right)^2 \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt) \\ =& \frac{1}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \frac{\partial \psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}} \frac{\partial \psi_0(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^T} \\ & (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \eta^2 \left(t^B - d \right)^2 \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt) \\ \le & \frac{\int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} c_2 \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \|_2^2 + \eta^2 (d - d_0)^2 \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt)}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2} \\ \le & \frac{\int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} c_2 \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \|_2^2 + \eta^2 \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \|_2^2 \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt)}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2} \\ \le & \frac{\int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} c_2 \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|_2^2 + \eta^2 \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|_2^2 \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt)}{n \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2}. \end{aligned}$$

Hence, there exists a positive constant c_3 in a neighbour of $\boldsymbol{\theta}_0$ such that

$$f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) \leq \frac{c_3}{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt),$$

and

$$\begin{aligned}
&\mathrm{E}\{f_1(\boldsymbol{\theta}; \boldsymbol{X}_i)\} \\
\leq &\frac{c_3}{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} \mathrm{E}\{R_i^B(dt^B; t) N_i(dt)\} \\
\leq &\frac{c_3 \mid d - d_0 \mid}{n} \\
\leq &\frac{c_3 \parallel \boldsymbol{\theta} - \boldsymbol{\theta}_0 \parallel}{n}.
\end{aligned}$$

Since

$$\sup_{\{\boldsymbol{\theta}: \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta\}} \mid f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) \mid \leq \frac{c_3}{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \leq t]} \mathbb{1}_{\left[t^B \in \mathcal{T}_2^B\right]} R_i^B(dt^B; t) N_i(dt),$$

we denote the envelope of $f_1(\boldsymbol{\theta}; \boldsymbol{X}_i)$, on the right side of the inequality, as F_{1i}, and it can be easily proved that $\mathrm{E}\{F_{1i}^2\} = O(1/n^2)$. Since $\{\boldsymbol{Z}, T\}$ has a positive and continuous density function (Condition (ii)), by Lemma 2.1 in Kim and Kim (2008)[13], there exists a positive constant c_4 such that

$$\begin{aligned}
&\mathrm{E}\left\{ \sup_{\{\boldsymbol{\theta}: \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \delta\}} \mid \sum_{i=1}^n \left(f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) - \mathrm{E} f_1(\boldsymbol{\theta}; \boldsymbol{X}_i)\right) \mid \right\} \\
\leq &c_4 \mathrm{E}\left\{ \sqrt{\sum_{i=1}^n F_{1i}^2} \right\} \\
\leq &c_4 \sqrt{\sum_{i=1}^n \mathrm{E} F_{1i}^2} \\
= &O(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Therefore,

$$\sum_{i=1}^n f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) \leq \sum_{i=1}^n \mathrm{E} f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) + \mid \sum_{i=1}^n f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) - \mathrm{E} f_1(\boldsymbol{\theta}; \boldsymbol{X}_i) \mid = O_p(\parallel \boldsymbol{\theta} - \boldsymbol{\theta}_0 \parallel) + O_p(1/\sqrt{n}),$$

and $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_2^B)^{(1)} = \parallel \boldsymbol{\theta} - \boldsymbol{\theta}_0 \parallel^2 \{\sum_{i=1}^n f_1(\boldsymbol{\theta}; \boldsymbol{X}_i)\} = o_p(\parallel \boldsymbol{\theta} - \boldsymbol{\theta}_0 \parallel^2)$.

Define

$$f_2(\boldsymbol{\theta};\boldsymbol{X}_i) = \frac{1}{n \parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel} \int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]} \left(Y_i^B(t^B;t) - \psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right) \left(\psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}) - \psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right) \mathbb{1}_{\left[t^B\in\mathcal{T}_2^B\right]} R_i^B(dt^B;t)N_i(dt),$$

and it is easily to prove that $\mathrm{E}\{f_2(\boldsymbol{\theta};\boldsymbol{X}_i)\} = 0$. Since $Y_i^B(t^B;t)$ and $\psi(t^B,t,\boldsymbol{Z}_i;\boldsymbol{\theta}_0)$ are bounded, there exists a positive constant c_5 such that

$$\sup_{\{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta\}} \mid f_2(\boldsymbol{\theta};\boldsymbol{X}_i) \mid \le \frac{c_5}{n}\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B\le t]}\mathbb{1}_{\left[t^B\in\mathcal{T}_2^B\right]} R_i^B(dt^B;t)N_i(dt).$$

We define the envelope of $f_2(\boldsymbol{\theta};\boldsymbol{X}_i)$, on the right side of the inequality, as F_{2i}, and it can be easily proved that $\mathrm{E}\{F_{2i}^2\} = O(\mid d-d_0 \mid /n)$. Again, by the Lemma 2.1 in Kim and Kim (2008), there exists a positive constant c_6 such that

$$\begin{aligned}
&\mathrm{E}\left\{\sup_{\{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta\}} \mid \sum_{i=1}^n \left(f_2(\boldsymbol{\theta};\boldsymbol{X}_i) - \mathrm{E}f_2(\boldsymbol{\theta};\boldsymbol{X}_i)\right) \mid\right\} \\
\le& c_6 \mathrm{E}\left\{\sqrt{\sum_{i=1}^n F_{2i}^2}\right\} \\
\le& c_6 \sqrt{\sum_{i=1}^n \mathrm{E}F_{2i}^2} \\
=& O(\frac{\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel}{\sqrt{n}}).
\end{aligned}$$

Therefore,

$$\sum_{i=1}^n f_1(\boldsymbol{\theta};\boldsymbol{X}_i) \le \sum_{i=1}^n \mathrm{E}f_2(\boldsymbol{\theta};\boldsymbol{X}_i) + \mid \sum_{i=1}^n f_2(\boldsymbol{\theta};\boldsymbol{X}_i) - \mathrm{E}f_2(\boldsymbol{\theta};\boldsymbol{X}_i) \mid = O(\frac{\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel}{\sqrt{n}}),$$

and $G_n(\boldsymbol{\theta},\boldsymbol{\theta}_0;\mathcal{T}_2^B)^{(2)} = \parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel \{\sum_{i=1}^n f_2(\boldsymbol{\theta};\boldsymbol{X}_i)\} = o_p(\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel /\sqrt{n})$.

Hence, we have that $G(\boldsymbol{\theta},\boldsymbol{\theta}_0;\mathcal{T}_2^B) = o_p(\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel^2) + o_p(\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel /\sqrt{n})$, and we can also prove that $G(\boldsymbol{\theta},\boldsymbol{\theta}_0;\mathcal{T}_3^B) = o_p(\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel^2) + o_p(\parallel \boldsymbol{\theta}-\boldsymbol{\theta}_0 \parallel /\sqrt{n})$ by similar arguments.

By applying Taylor Expansion to $G(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_1^B)$ and $G(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_4^B)$, we have

$$
\begin{aligned}
& G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_1^B) + G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0; \mathcal{T}_4^B) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left[\left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|) \right\}^2 - \right. \\
& \left. 2 \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|) \right\} \right] \\
& \mathbb{1}_{\left[t^B \in \mathcal{T}_1^B\right]} R_i^B(dt^B; t) N_i(dt) \\
& + \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left[\left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|) \right\}^2 - \right. \\
& \left. 2 \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|) \right\} \right] \\
& \mathbb{1}_{\left[t^B \in \mathcal{T}_4^B\right]} R_i^B(dt^B; t) N_i(dt) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left[(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \left(\boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) A_1(t^B, t; \boldsymbol{\theta}_0)^T \mathbb{1}_{[t^B \le d_0]} \right. \right. \\
& \left. + \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0)^T \mathbb{1}_{[t^B > d_0]} \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
& - 2 \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \left(\boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B \le d_0]} + \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B > d_0]} \right) \\
& \left. + \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|) \right] R_i^B(dt^B; t) N_i(dt) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \\
=& (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \left[\frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left\{ \boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) \boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0)^T \mathbb{1}_{[t^B \le d_0]} + \right. \right. \\
& \left. \left. \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0)^T \mathbb{1}_{[t^B > d_0]} \right\} R_i^B(dt^B; t) N_i(dt) \right] (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\
& - \frac{2 (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T}{\sqrt{n}} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \right. \\
& \left. \left(\boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B \le d_0]} + \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B > d_0]} \right) R_i^B(dt^B; t) N_i(dt) \right] \\
& + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| / \sqrt{n}).
\end{aligned}
$$

Define

$$
\begin{aligned}
\boldsymbol{W}_n =& \frac{1}{n} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left\{ \boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) \boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0)^T \mathbb{1}_{[t^B \le d_0]} + \right. \\
& \left. \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0)^T \mathbb{1}_{[t^B > d_0]} \right\} R_i^B(dt^B; t) N_i(dt),
\end{aligned}
$$

and by the Law of Large Number $\boldsymbol{W}_n = \boldsymbol{W} + o_p(1)$ where $\boldsymbol{W} = \mathrm{E}\{W_n\}$. Define

$$\begin{aligned}\boldsymbol{V}_n =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \left(\boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B \le d_0]} + \right. \\ & \left. \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B > d_0]} \right) R_i^B(dt^B; t) N_i(dt),\end{aligned}$$

and by the Central Limit Theorem $\boldsymbol{V}_n$ converge weakly to a mean zero multivariate Gaussian variable $\boldsymbol{V}$ i.e. $\boldsymbol{V} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and

$$\begin{aligned}\boldsymbol{\Sigma} =& \mathrm{Var} \left\{ \int_0^{+\infty} \int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left(Y_i^B(t^B; t) - \psi(t^B, t, \boldsymbol{Z}_i; \boldsymbol{\theta}_0) \right) \left(\boldsymbol{A}_1(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B \le d_0]} + \right. \right. \\ & \left. \left. \boldsymbol{A}_2(t^B, t; \boldsymbol{\theta}_0) \mathbb{1}_{[t^B > d_0]} \right) R_i^B(dt^B; t) N_i(dt) \right\}.\end{aligned}$$

Therefore,

$$\begin{aligned}& G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \\ =& (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{W} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{2 (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T}{\sqrt{n}} \boldsymbol{V}_n + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| / \sqrt{n}) \\ =& \left\{ \boldsymbol{W}^{\frac{1}{2}} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{\boldsymbol{W}^{-\frac{1}{2}} \boldsymbol{V}_n}{\sqrt{n}} \right\}^T \left\{ \boldsymbol{W}^{\frac{1}{2}} (\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{\boldsymbol{W}^{-\frac{1}{2}} \boldsymbol{V}_n}{\sqrt{n}} \right\} - \frac{\boldsymbol{V}_n^T \boldsymbol{W}^{-1} \boldsymbol{V}_n}{n} \\ & + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|^2) + o_p(\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| / \sqrt{n}).\end{aligned}$$

Consider the convergence rate in the Theory (4.3.2) and the fact that $\hat{\boldsymbol{\theta}}$ reaches the minimum point of $Q_n(\boldsymbol{\theta})$, we have

$$G_n(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}_0) \le G_n(\boldsymbol{\theta}_0 + \frac{\boldsymbol{W}^{-1} \boldsymbol{V}_n}{\sqrt{n}}, \boldsymbol{\theta}_0) = \frac{\boldsymbol{V}_n^T \boldsymbol{W}^{-1} \boldsymbol{V}_n}{n} + o_p(\frac{1}{n}).$$

Since $\hat{\boldsymbol{\theta}}$ is consistent and $G_n(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ is continuous in terms of $\boldsymbol{\theta}$, we have $\| \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \frac{\boldsymbol{W}^{-1} \boldsymbol{V}_n}{\sqrt{n}} \|^2 = o_p(1/n)$, and therefore $\sqrt{n} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) = \boldsymbol{W}^{-1} \boldsymbol{V}_n + o_p(1) \to \boldsymbol{W}^{-1} \boldsymbol{V} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{W}^{-1} \boldsymbol{\Sigma} \boldsymbol{W}^{-1})$. So the desired result is proved.

4.7.5 Proof of Theorem (4.3.5)

We have

$$
\begin{aligned}
&\mathcal{Q}(\boldsymbol{\theta}) - \mathcal{Q}(\boldsymbol{\theta}_0) \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left(\psi(t^B, t, \boldsymbol{Z}; \boldsymbol{\theta}) - \psi(t^B, t, \boldsymbol{Z}; \boldsymbol{\theta}_0)\right)^2 \mathrm{E}\left\{R^B(dt^B; t) | N(dt), \boldsymbol{Z}\right\}\right. \\
&\left.\mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right] \\
=&\mathrm{E}\left[\int_0^{+\infty}\int_0^{+\infty} \mathbb{1}_{[t^B \le t]} \left(\psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}) - \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}_0)\right)^2 \mathbb{1}_{[t^B \le d \wedge d_0]}\right. \\
&+ \mathbb{1}_{[t^B \le t]} \left(\psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}) + \eta_1 + \eta_2(t^B - d) - \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}_0)\right)^2 \mathbb{1}_{[d < t^B \le d_0]} \\
&+ \mathbb{1}_{[t^B \le t]} \left(\psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}) - \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}_0) - \eta_{10} - \eta_{20}(t^B - d_0)\right)^2 \mathbb{1}_{[d_0 < t^B \le d]} \\
&+ \mathbb{1}_{[t^B \le t]} \left(\psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}) + \eta_1 + \eta_2(t^B - d) - \psi_0(t^B, t, \boldsymbol{Z}; \boldsymbol{\beta}) - \eta_{10} - \eta_{20}(t^B - d_0)\right)^2 \\
&\left.\mathbb{1}_{[t^B > d \vee d_0]} \mathrm{E}\left\{R^B(dt^B; t) | N(dt) = 1, \boldsymbol{Z}\right\} \mathrm{E}\left\{N(dt)|\boldsymbol{Z}\right\}\right]
\end{aligned}
\tag{4.17}
$$

We denote the four terms in the above equation as $B1 - 4$. By using the similar arguments in the proof of Theorem (4.3.2), on can prove that based on the conditions (iv* -vii*) there exist positive constants c_7 such that in a small neighbourhood of $\boldsymbol{\theta}_0$

$$
B_1 + B4 \ge c_7 \left(\| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \|_2^2 + \| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \|_2^2\right) \tag{4.18}
$$

Because $\eta_{10} \neq 0$, one can easily prove that there exist positive constants c_8 such that in a small neighbourhood of $\boldsymbol{\theta}_0$ $B_2 + B_3 \ge c_8 |d - d_0|$. Therefore, one can prove that there exist positive constants c_9 such that in a small neighbourhood of $\boldsymbol{\theta}_0$

$$
\mathcal{Q}(\boldsymbol{\theta}) - \mathcal{Q}(\boldsymbol{\theta}_0) \ge c_9 \rho_2(\boldsymbol{\theta}, \boldsymbol{\theta}_0)^2. \tag{4.19}
$$

Applying the condition (v) again, we have the desired convergence rate(Kosorok, Theorem 14.4, 2008).

4.7.6 Proof of Theorem (4.3.6)

Since conditions (i-iii) guarantee that $\hat{\boldsymbol{\theta}}$ is consistent and $\boldsymbol{\theta}_0$ is assumed to line in the interior of $\boldsymbol{\Theta}$, we know that $\hat{\boldsymbol{\theta}}$ lies in a small neighborhood of $\boldsymbol{\theta}_0$ and cannot be

the boundary with sufficiently large probability. This implies that the maximum is a local maximum and as a result $\partial Q_n(\hat{d}, \hat{\boldsymbol{\gamma}})/\partial \boldsymbol{\gamma} = \mathbf{0}$. By Mean Value Theorem, we have

$$\frac{\partial Q_n(\hat{d}, \hat{\boldsymbol{\gamma}})}{\partial \boldsymbol{\gamma}} = \frac{\partial Q_n(\hat{d}, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma}} + \frac{\partial^2 Q_n(\hat{d}, \boldsymbol{\gamma}^*)}{\partial \boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0), \tag{4.20}$$

where $\| \boldsymbol{\gamma}^* - \boldsymbol{\gamma}_0 \|_2 \leq \| \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \|_2$. We define $J_n(d, \boldsymbol{\gamma}) = -\partial^2 Q_n(d, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T$. By the Law of Large Number, $J_n(d, \boldsymbol{\gamma}) \to \boldsymbol{I}_1(\boldsymbol{\theta})$ Thus, we have

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \sqrt{n} J_n^{-1}(\hat{d}, \boldsymbol{\gamma}^*) \frac{\partial Q_n(\hat{d}, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} = \sqrt{n} J_n^{-1}(d_0, \boldsymbol{\gamma}^*) \frac{\partial Q_n(d_0, \boldsymbol{\gamma}_0)}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T} + \sqrt{n} O(|d_0 - \hat{d}|). \tag{4.21}$$

The second equation is because $J_n^{-1}(d, \boldsymbol{\gamma})$ and $\partial Q_n(\hat{d}, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T$ are uniformly continuous with respect to d. Theorem (4.3.5) implies that $\sqrt{n} O(|d_0 - \hat{d}|) = o(1)$. Since $\hat{\boldsymbol{\gamma}}$ is consistent and by condition (vi), $J_n^{-1}(d_0, \boldsymbol{\gamma}^*)$ converges to $\boldsymbol{I}_1(\theta_0)^{-1}$ in a small neighborhood of $\boldsymbol{\theta}_0$. By Central Limit Theorem, $\sqrt{n} \partial Q_n(d_0, \boldsymbol{\gamma}_0)/\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^T \to N(\mathbf{0}, \boldsymbol{I}_2(\theta_0))$ weakly. By Slutsky's Theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \to N\left(\mathbf{0}, \boldsymbol{I}_1(\boldsymbol{\theta}_0)^{-1} \boldsymbol{I}_2(\boldsymbol{\theta}_0) \boldsymbol{I}_1(\boldsymbol{\theta}_0)^{-1}\right) \tag{4.22}$$

weakly in a small neighborhood of $\boldsymbol{\theta}_0$.

Chapter 5

Disccusion and Future Research

In this dissertation, we have focused on the approaches for joint modeling of longitudinal measurements, recurrent events, and failure time data which are frequently observed in medical studies. As introduced in Chapter 1, there are two types of longitudinal measurements: (i) repeated measurements collected at sampling times, and (ii) marker measurements observed when recurrent events occur. In practice, the longitudinal measurments, recurrent events, and failure time data are usually correlated with each other for both types of data. Ignoring the correlation within the data structure will lead to non-negligible biases and therefore the traditional methods for longitudinal data analysis (Liang and Zeger, 1986[19]; Laird and Ware, 1982[18]) can not be applied in the situation where sampling times and failure times are dependent with the longitudinal measurements.

We have developed a new joint modeling method for longitudinal measurements, recurrent events, and failure events data where all of the three components of data are treated as outcomes. Without requiring restrict assumptions on recurrent event processes, the proposed model is applicable to both types of longitudinal measurements, (i) and (ii). On the basis of the survivors' population, the model avoids the disputing assumption on the existence of recurrent events or longitudinal measurements after the failure event. As the model does not involve latent variables, computationally it is simpler and easier to adopt when comparing to the shared frailty

model. The proposed model and estimation inference can be generalized to analyze left-truncated and right-censored data. The proposed model involves semiparametric structure in each of the three sub-models, for failure time, recurrent events and longitudinal measurements, and the baseline functions in the model are unspecified. Our model also possesses a specific feature that the forward time model is equivalent to the backward-in-time model for recurrent events and longitudinal measurements, where the two models share the same regression parameter values.

In addition, it is also of interest to study the terminal behavior of biomarkers prior to the occurrence of failure events by aligning failure events as time origins and counting time backward. This research topic is primarily motivated by the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) cohort study which aims at identifying biomarkers associated with the development of Alzheimer's Disease (AD) procession. In studies of Alzheimer's Disease the rate of change in biomarker measurement before diagnosis of disease is widely recognized as an important index for predicting the disease (Hall et al., 2000[8]; Wilson et al., 2007[43]). Moreover, when studying trajectories of biomarkers or other longitudinal measurements before the failure event, researchers sometimes find the phenomenon that the rate of change of biomarkers would start to shift, accelerate or decelerate, at some special time point prior to the failure event by a gap time which is named as the backward change point. However, the conventional forward perspective of stochastic processes are not designed for the terminal behavior of processes. It is even more challenging to estimate the backward change point because of the censoring problem and the complex data structure.

In this dissertation, we have proposed a statistical method to study the terminal behavior of stochastic processes and estiamte the backward change point prior to the failure event. Our approach cleverly solves the censoring problem which is, to our knowledge, usually ignored when estimating the backward change point. Especially,

our method allows the censoring time to be associated with the failure time and therefore can be applied with the occurence of competing risks issue. With the spirit of semiparametric models, our method has few restrictions on the variable distributions or the correlations within the data structure except for a few common assumptions. We also studied the asymptotic properties of our method and confirmed them by simulation studies. It is noticed that we have only provied the asymptotic distribution for the continuous backward change point estimator in this dissertation and developing the asymptotic distribuition of the sudden-jump backward change point estimator is more difficult and needs further study. We have illustrated our method by applying it to the BIOCARD data and the applications on other data sets can be done in the future.

There are still many open research topics not addressed in this dissertation. First, the joint modeling method in Chapter 3 only has only considered the longitudinal measurements, recurrent events, and failure time data under the assumption of independent censoring, and therefore can not be applied in the situation where competing risk occurs. Competing risk is an event that either avoids the observation of the failure event of interest or changes the opportunity of the occurence of the failure event. Competing risk is commonly observed in survival analysis. For example, in studies of Alzheimer's Disease, subjects can die due to cancer or other diseases before the onset of clinical symptoms of Alzheimer's Disease and the death will terminate the observation of the development of Alzheimer's Disease and corresponding longitudinal measurements. The competing risk is usually correlated with the failure event of interest and therefore models which are developed based on the assumption of independent censoring can not be applied to analyze data with competing risks. Joint modeling of longitudinal measurements, recurrent events, and failure time data with competing riks is of practical meaning but very challenging, and therefore more efforts can be made in this direction for future study.

Second, the backward change point models in Chapter 4 take few advantage of the available data of the censored subjects. How to use the data from the censored subjects can be an interesting research topic in the future because the censoring rate can be very high in many follow up studies (e.g. the censoring rate is about 0.8 in Biocard cohort study). Moreover, there have been many studies in developing forward-in-time models for stochastic processes and some researchers have started to consider models in backward time scale, but few efforts have been conducted to combine models in forward and backward together. We think that modeling in both forward and backward time index may be a solution to using the data of censored subjects and should be considered as the future work.

In conclusion, the proposed joint models for longitudinal measurements, recurrent events, and failuretime data and the backward change point models in this dissertation may initiate a wave of future research on both statistical applications and methods. The proposed models can help to answer a wide range of public health or medical scientific questions and understand the nature history of biology processes.

Bibliography

[1] Donald I Abrams, Anne I Goldman, Cynthia Launer, Joyce A Korvick, James D Neaton, Lawrence R Crane, Michael Grodesky, Steven Wakefield, Katherine Muth, Sandra Kornegay, et al. A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, 330(10):657–662, 1994.

[2] Ivan SF Chan, James D Neaton, Louis D Saravolatz, Lawrence R Crane, and James Osterberger. Frequencies of opportunistic diseases prior to death among hiv-infected persons. *Aids*, 9(10):1145–1152, 1995.

[3] Kwun Chuen Gary Chan and Mei-Cheng Wang. Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events. *Journal of the American Statistical Association*, 0(ja):1–31, 0.

[4] Kwun Chuen Gary Chan and Mei-Cheng Wang. Backward estimation of stochastic processes with failure events as time origins. *The annals of applied statistics*, 4(3):1602, 2010.

[5] LA Farrer, LA Cupples, JL Haines, B Hyman, WA Kukull, R Mayeux, RH Myers, MA Pericak-Vance, N Risch, and CM Van Duijn. Apoe and alzheimer disease meta analysis consortium: Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease: a meta-analysis. *Jama*, 278(16):1349–1356, 1997.

[6] Lindsay A Farrer, L Adrienne Cupples, Jonathan L Haines, Bradley Hyman, Walter A Kukull, Richard Mayeux, Richard H Myers, Margaret A Pericak-Vance, Neil Risch, and Cornelia M Van Duijn. Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease: a meta-analysis. *Jama*, 278(16):1349–1356, 1997.

[7] Debashis Ghosh and DY Lin. Semiparametric analysis of recurrent events data in the presence of dependent censoring. *Biometrics*, 59(4):877–885, 2003.

[8] Charles B Hall, Richard B Lipton, Martin Sliwinski, and Walter F Stewart. A change point model for estimating the onset of cognitive decline in preclinical alzheimer's disease. *Statistics in medicine*, 19(11-12):1555–1566, 2000.

[9] Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

[10] Chiung-Yu Huang and Mei-Cheng Wang. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*, 99(468):1153–1165, 2004.

[11] Yijian Huang and Mei-Cheng Wang. Frequency of recurrent events at failure time: modeling and inference. *Journal of the American Statistical Association*, 98(463):663–670, 2003.

[12] Clifford R Jack, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[13] Jeankyung Kim and Hyune-Ju Kim. Asymptotic results in segmented multiple regression. *Journal of Multivariate Analysis*, 99(9):2016–2038, 2008.

[14] Sehee Kim, Donglin Zeng, Lloyd Chambless, and Yi Li. Joint models of longitudinal data and recurrent events with informative terminal event. *Statistics in biosciences*, 4(2):262–281, 2012.

[15] Michael R Kosorok. Introduction to empirical processes. *Introduction to Empirical Processes and Semiparametric Inference*, pages 77–79, 2008.

[16] Michael R Kosorok, Rui Song, et al. Inference under right censoring for transformation models with a change-point based on a covariate threshold. *The Annals of Statistics*, 35(3):957–989, 2007.

[17] MR Kosorok. Introduction to empirical processes and semiparametric inference. 2008.

[18] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[19] Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

[20] Yu Liang, Wenbin Lu, and Zhiliang Ying. Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65(2):377–384, 2009.

[21] DY Lin, LJ Wei, I Yang, and Z Ying. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730, 2000.

[22] DY Lin and Z Ying. Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–126, 2001.

[23] Haiqun Lin, Daniel O Scharfstein, and Robert A Rosenheck. Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):791–813, 2004.

[24] Lei Liu and Xuelin Huang. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):65–81, 2009.

[25] Lei Liu, Xuelin Huang, and John O'Quigley. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics*, 64(3):950–958, 2008.

[26] Lei Liu, Robert A Wolfe, and Xuelin Huang. Shared frailty models for recurrent events and a terminal event. *Biometrics*, 60(3):747–756, 2004.

[27] June R Lunney, Joanne Lynn, Daniel J Foley, Steven Lipson, and Jack M Guralnik. Patterns of functional decline at the end of life. *Jama*, 289(18):2387–2392, 2003.

[28] James D Neaton, Deborah N Wentworth, Frank Rhame, Carlton Hogan, Donald I Abrams, and Lawrence Deyton. Considerations in choice of a clinical endpoint for aids clinical trials. *Statistics in medicine*, 13(19-20):2107–2125, 1994.

[29] Odile Pons et al. Estimation in a cox regression model with a change-point according to a threshold in a covariate. *The Annals of Statistics*, 31(2):442–463, 2003.

[30] Henrik Ramlau-Hansen. Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, pages 453–466, 1983.

[31] Steven J Skates, Donna K Pauler, and Ian J Jacobs. Screening based on the risk of cancer calculation from bayesian hierarchical changepoint and mixture models of longitudinal markers. *Journal of the American Statistical Association*, 96(454):429–439, 2001.

[32] Elizabeth H Slate and Kathleen A Cronin. Changepoint modeling of longitudinal psa as a biomarker for prostate cancer. In *Case studies in Bayesian statistics*, pages 435–456. Springer, 1997.

[33] Xiao Song, Marie Davidian, and Anastasios A Tsiatis. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58(4):742–753, 2002.

[34] Xiao Song and CY Wang. Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics*, 64(2):557–566, 2008.

[35] Jianguo Sun, Do-Hwan Park, Liuquan Sun, and Xingqiu Zhao. Semiparametric regression analysis of longitudinal data with informative observation times. *Journal of the American Statistical Association*, 100(471):882–889, 2005.

[36] Jianguo Sun, Liuquan Sun, and Dandan Liu. Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association*, 102(480):1397–1406, 2007.

[37] Liuquan Sun, Xinyuan Song, Jie Zhou, and Lei Liu. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107(498):688–700, 2012.

[38] Len A Usvyat, Claudia Barth, Inga Bayh, Michael Etter, Gero D von Gersdorff, Aileen Grassmann, Adrian M Guinsburg, Maggie Lam, Daniele Marcelli, Cristina

Marelli, et al. Interdialytic weight gain, systolic blood pressure, serum albumin, and c-reactive protein levels change in chronic dialysis patients prior to death. *Kidney international*, 84(1):149–157, 2013.

[39] Aad W Van Der Vaart and Jon A Wellner. Weak convergence. In *Weak Convergence and Empirical Processes*, pages 16–28. Springer, 1996.

[40] Edward F Vonesh, Tom Greene, and Mark D Schluchter. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in medicine*, 25(1):143–163, 2006.

[41] Mei-Cheng Wang, Ron Brookmeyer, and Nicholas P Jewell. Statistical models for prevalent cohort data. *Biometrics*, pages 1–11, 1993.

[42] Mei-Cheng Wang, Jing Qin, and Chin-Tsang Chiang. Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, 96(455):1057–1065, 2001.

[43] Robert S Wilson, Todd L Beck, Julia L Bienias, and David A Bennett. Terminal cognitive decline: accelerated loss of cognition in the last years of life. *Psychosomatic Medicine*, 69(2):131–137, 2007.

[44] Robert S Wilson, Sue E Leurgans, Patricia A Boyle, and David A Bennett. Cognitive decline in prodromal alzheimer disease and mild cognitive impairment. *Archives of neurology*, 68(3):351–356, 2011.

[45] Michael S Wulfsohn and Anastasios A Tsiatis. A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339, 1997.

[46] Hua Xiang, Mei-Cheng Wang, and Chiung-Yu Huang. A comparison of various rate functions of a recurrent event process in the presence of aterminal event. *Statistical Methods in Medical Research*, 2008.

[47] Jane Xu and Scott L Zeger. Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):375–387, 2001.

[48] Yining Ye, John D Kalbfleisch, and Douglas E Schaubel. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*, 63(1):78–87, 2007.

[49] Zhiliang Ying. A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, pages 76–99, 1993.

[50] Donglin Zeng and Jianwen Cai. A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika*, page asq039, 2010.

Qing Cai

615 North Wolfe Street, E3033
Baltimore, MD, 21205, USA

Phone: 443-310-4968
Email: qcai3@jhu.edu

Education

- B.S. Mathematics and Statistics, Nanjing University, June 2012.
- Ph.D. Biostatistics, Johns Hopkins University, May 2017 (Expected).

Research Interests

Survival analysis; Longitudinal and biomarker data analysis; Recurrent events data; Backward stochastic processes; Change point detection

Professional Experience

Research Assistant

[2013-present] Worked with Dr. Mei-Cheng Wang, Dr. Marilyn Albert and other investigators in the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD) cohort study which aims at identifying biomarkers associated with the development of Alzheimer's Disease (AD) procession, and the selected projects inluded:

- Analyzed the predictive ability of the proposed clinical index scores which were combinations of clinical symptoms, demographic variables, and neuropsychological test results for the onset of neuropsychological test results.
- Examined the Preclinical AD Stages determined by CSF levels in relation to longitudinal changes in MRI biomarker levels over time.
- Determined if the rate of change in cognitive performance over time was associated with the baseline level of cognitive reserve and the baseline levels of

AD-pathology, as measured by a composite score.

Teaching Assistant

[2013-present] Assisted professors to lead lab sessions, hold office hours, grade homework and exams for the following courses:

2013 Fall	Survival Analysis I-II
2014 Spring	Statistics for Laboratory Scientists I-II
2014 Fall	Survival Analysis I-II
2015 Spring	Statistical Methods in Public Health III-IV
2015 Fall	Survival Analysis I-II
2016 Spring	Biostatistics for Laboratory Scientists I-II
2016 Fall	Survival Analysis I, Statistical Methods in Public Health II

Honors & Awards

2016	First place in the Student Paper Award Competition of the Mental Health Section of the ASA
2011	Samsung Scholarship, Nanjing University

Papers

Published

- Resnick, S. M., Bilgel, M., Moghekar, A., An, Y., Cai, Q., Wang, M. C., ... & Wong, D. F. (2015). Changes in Aβ biomarkers and associations with APOE genotype in 2 longitudinal cohorts. Neurobiology of Aging, 36(8), 2333-2339.

- Soldan, A., Pettigrew, C., Cai, Q., Wang, M. C., Moghekar, A. R., OBrien, R. J., ... & Albert, M. S. (2016). Hypothetical preclinical Alzheimer disease groups and longitudinal cognitive change. JAMA Neurology, 73(6), 698-705.

- Cai, Q., Wang, M.C., & Chan, K.C.G. (2017) Joint modeling of longitudinal, recurrent events and failure time data for survivor's population. Biometrics, In Press.

Manuscripts in Preparation

- Cai, Q. & Wang, M.C. Change point estimation in backward process model.
- Sacktor, N., Soldan, A., Grega, M., Farrington, L., Cai, Q., Wang, M-C., Gottesman, R., Turner, R., & Albert, M. The BIOCARD Index: A Summary Measure to Predict Onset of Mild Cognitive Impairment.
- Soldan, A., Pettigrew, C., Cai. Q., Wang, M-C., Moghekar, A., Miller, M.I., & Albert, M. Association between cognitive reserve and cognitive change during preclinical Alzheimer's disease.

Conference Abstracts

- Soldan, A., Pettigrew, C., Cai, Q., Wang, M.C., Moghekar, A., Brown, T., Miller, M.I., & Albert, M. (2017). Cognitive reserve and longitudinal trajectories of AD biomarkers. Session on Cognitive Reserve and Compensatory Processes, Dallas Aging and Cognition Conference, Dallas, TX.
- Sacktor, N., Soldan, A., Grega, M., Farrington, L., Cai, Q., Wang, Mei-Cheng, Gottesman, R. F., & Albert, M. (2017). The BIOARD Index: A summary measure to predict onset of Mild Cognitive Impairment. American Academy of Neurology annual meeting, Boston, MA.
- Soldan, A., Pettigrew, C., Cai, Q., Wang, M.C., & Albert, M. (2016). Cognitive reserve and longitudinal cognitive change before and after onset of cognitive impairment. Data Blitz Session, Professional Interest Area: Resilience, Resilience, and Protective Factors, Alzheimer Association International Conference, Toronto, Canada.
- Soldan, A., Pettigrew, C., Cai, Q., Wang, M.C., & Albert, M. (2016). Relationship between cognitive reserve and longitudinal change in cognition in

middle-aged and older adults. Plenary Session: Mild Cognitive Impairment. Cognitive Aging Conference, Atlanta, GA.

- Soldan, A., Pettigrew, C., Cai, Q., Wang, M.C., & Albert, M. (2016). Relationship between cognitive reserve and longitudinal change in cognition in middle-aged and older adults. Symposium on Resilience to Brain Aging and Alzheimer's Disease: Evidence from Imaging and Biomarker Studies. International Neuropsychological Society Meeting, Boston, MA.

- Soldan, A., Pettigrew, C., Cai, Q., Wang, M.C., Moghekar, A., Miller, M.I., Ratnanather, T., Mori, S., & Albert, M.S. (2015). Relationship of CSF Tau and β-Amyloid to Hippocampal Atrophy Rates. Poster presentation, Alzheimer's Association International Conference, Washington, DC.